Studying the principles of cell identity transitions using naïve pluripotency induction as a model

Hannah Taylor Stuart

This dissertation is submitted for the degree of
Doctor of Philosophy

Supervised by José Silva, PhD

Gonville & Caius College
University of Cambridge

June 2018

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically stated in the text.

It does not exceed the word limit for the Degree Committee of the Faculty of Biology.

Hannah Taylor Stuart

Studying the principles of cell identity transitions using naïve pluripotency induction as a model

Hannah Taylor Stuart

Summary

In multicellular biology, an astounding array of cellular identities are specified from the same genome, by drawing on a finite pool of transcription factors and signalling pathways in different permutations. How are signal and transcription network interplay computed by the cell to instruct identity? Must a given identity always be established by the same mechanism?

To address these questions, I created defined and tractable systems based on reprogramming from epiblast stem cells to naïve pluripotency. By independent modulation of genetic and signalling parameters followed by isolation of successfully transitioning cells, I showed that naïve pluripotency can be established via distinct routes. These differ in their transcriptional trajectories and in their mechanisms, each with different genetic and signal requirements. Relative to development, one route initially moves forward, with productive cells acquiring mesodermal signature prior to naïve pluripotency induction. In contrast, another route overshoots backwards, transiently resembling the earlier embryo and recapitulating key aspects of naïve epiblast establishment *in vivo.*

These remarkably distinct routes ultimately converge on the same naïve pluripotent endpoint, revealing surprising flexibility for the establishment of a single identity from a single origin. This provides evidence for cellular identity as a multidimensional attractor state (Kauffman, 1993), and extends the paradigm by which transcription factors and signals are used in different combinations to generate different cell types: they can also be used in different ways to generate the same cell type.

How can the naïve pluripotent identity be reached via such different routes? I reconciled route differences, finding precise Oct4 expression as a unifying, essential and sufficient feature. Different routes achieve this required Oct4 expression by diverse logics, highlighting that successful identity change does not simply require activation of the destination network but instead pivots on the mechanism that permits a transition to occur.

Fixing Oct4 at this level is sufficient to reprogram EpiSCs and more advanced cell types under signal instruction alone, yet is also compatible with bonafide development when returned to the *in vivo* context. At this expression level, I therefore deem Oct4 to be a ‘transition’ factor, a novel concept that I define as permitting identity change in various directions depending on environmental instruction.

In sum, this thesis furthers our knowledge of naïve pluripotency induction *in vitro* and *in vivo*, and provides conceptual advances in our understanding of cell identity transitions.

Acknowledgements

I thank José Silva for the opportunity to pursue a PhD in his laboratory, for his inexorable optimism, his unwavering support and patience during difficult times, his instinctive ability to see the wood for the trees, and above all for giving me the trust and freedom to develop independently as a scientist. I have learned so much both professionally and personally throughout these years, and am truly lucky to move forward in my career from this springboard.

Working together with Jenny Nichols profoundly enriched both the project and my doctoral experience. In addition to her wealth of expertise and invaluable collaboration, I am deeply grateful for her voluntary adoption of a mentorship role towards me. Our many discussions are a highlight of my PhD. I thank her for always being there, and for continually expanding my horizons. Her scientific rigour, enthusiasm and determination are inspirational.

I am grateful to Bertie Göttgens and his team members Sonia Nestorowa and Chee Lim for their generous and professional collaboration on the single-cell RNA sequencing experiments. Likewise, I am indebted to Giuliano Stirparo, Paul Bertone and Sabine Dietmann for their vital work with the sequencing and microarray datasets. In particular, thank you to Giuliano for his patience and solidarity through all those days of analysis together.

Thank you to Austin Smith for his feedback and sharp insight. Although infrequent, our discussions proved formative to my way of thinking, and capacitated my ongoing work. Many members of the Smith group also provided me with essential training and advice throughout my PhD. In particular, I would like to thank Sam Myers and James Clarke for teaching me how to culture human ESCs, Masaki Kinoshita for assistance analysing mid-gestation chimeras, and Tüzer Kalkan and Carla Mulas for their willingness to share their encyclopedic knowledge.

Over these years, I have been fortunate to supervise a number of excellent and entertaining undergraduate, masters and PhD students. Mariana Alves, Petros Fessas, Becca Lloyd, Tim Lohoff, Kate Maskalenska and Andrew Malcolm were a pleasure to have in the team, and their contributions were instrumental for the research progression.

I thank all members of the Silva lab for supporting me, occasionally literally. The sense of camaraderie amongst this team of peers is unique, and I have made friends for life here. Lawrence Bates, the pillar of the lab, thank you for making everything work. Charlotte Handford, you are more glorious than soup. Aliaksandra Radzisheuskaya, your discoveries laid the foundations. Rodrigo Santos, for your perspective. Elsa Sousa, I'm so glad you eXist.

This work would not have been possible without specialist technical support from the facilities at the Stem Cell Institute and the Cambridge Institute for Medical Research. In particular, I would like to thank Kate Bird, Chiara Cossetti, Charles-Etienne Dumeau, Peter Humphreys, William Mansfield, Sally Lees, Vicki Murray, Maike Paramor, Irina Pshenichnaya and Andy Riddell. Not only have they offered consistent excellence, the attitude with which they rallied for manuscript revisions turned a stressful period into a fun and rewarding one.

I am grateful to the Medical Research Council for funding my studentship, and for their generosity and understanding in the face of health complications. Likewise, I would like to thank Dr Gall, Prof Khullar, Dr Lulsegged and Prof Seneviratne for their outstanding care. Together with my health, they restored my hopes of pursuing a scientific research career.

Thank you to my family and close friends. Your love and loyalty mean the world to me. Grraziano, for your fortitude and humoor, and for giving me so much to look forward to.

I would like to dedicate this thesis to my parents and grandparents. The values they instilled and their support of my education made this journey possible. Gran, I wish I could let you know that I finally finished, and that those little cells I worked with didn't jump out and bite me.

Contents

Summary i
Acknowledgements ii
Contents iv
List of Figures viii
List of Tables ix
List of Common Abbreviations x
CHAPTER 1: GENERAL INTRODUCTION 1
1.1 Cellular identity 1
1.1.1 What is cellular identity? 1
1.1.2 What determines cellular identity? 1
1.1.3 Identity stability 2
1.1.4 Cell identity transitions 3
1.1.5 Identity potency 5
1.2 Pluripotency 7
1.2.1 Naïve pluripotency capture and culture 7
1.2.2 The transcription factor network governing naïve pluripotency 10
1.2.3 Interplay between signal inputs and transcriptional control of naïve pluripotency 13
1.2.4 Epigenetic landscape of naïve pluripotency 15
1.2.5 Generation of naïve pluripotency by nuclear reprogramming 16
1.2.6 Primed pluripotency 17
1.2.7 Human pluripotency 21
1.3 Experimental study of cell identity transitions 22
1.3.1 Outstanding questions regarding mechanisms of cell identity transition 22
1.3.2 Requirements of an appropriate identity transition model 23
1.3.3 Choice of naïve pluripotency as the destination cell identity 23
1.3.4 Choice of EpiSC reprogramming to naïve pluripotency as the model system 24
1.3.5 Challenges to overcome 24
1.3.6 Summary of aims 26
CHAPTER 2: MATERIALS AND METHODS 27
2.1 Tissue culture 27
2.1.1 Murine tissue culture 27

2.1.2 Murine cell transfection 28
2.1.3 EpiSC reprogramming 28
2.1.4 FixedOct4 EpiSCs 29
2.1.5 Somatic cell reprogramming from E9.5 chimeras 29
2.1.6 Human tissue culture 29
2.1.7 Human cell transfection 29
2.1.8 Human cell resetting 30
2.2 Embryo work 30
2.2.1 Derivation of EpiSCs 30
2.2.2 Embryo culture and analysis 31
2.2.3 Blastocyst injection 31
2.3 Flow cytometry 32
2.4 Immunohistochemistry 32
2.5 Histology 32
2.6 Western blotting 33
2.7 Genotyping 34
2.7.1 Nanog genotyping 34
2.7.2 Oct4 genotyping 34
2.8 Plasmids and cloning 35
2.9 RT-qPCR 35
2.10 ChIP 37
2.10.1 ChIP procedure 37
2.10.2 Analysis of published ChIP-seq datasets 38
2.11 Microarray 38
2.12 RNAseq 38
2.12.1 Bulk RNAseq library preparation 38
2.12.2 scRNAseq library preparation 39
2.12.3 RNAseq alignment 39
2.12.4 Published embryo scRNAseq datasets 39
2.12.5 Transcriptome analysis 39
2.12.6 Selection of high-variability genes 40
2.12.7 Quadratic programming 40
2.12.8 LOESS regression 40

CHAPTER 3: NANOG AMPLIFIES STAT3 ACTIVATION AND THEY SYNERGISTICALLY INDUCE THE NAÏVE PLURIPOTENT PROGRAM 41
3.1 Introduction 41
3.2 Results 43
3.2.1 Nanog amplifies Stat3 phosphorylation and suppresses Socs3 transcription 43
3.2.2 Nanog and PStat3 synergistically upregulate Klf4 and novel *Klf4* enhancer RNA 47
3.2.3 Nanog and PStat3 induce rapid and efficient reactivation of naïve genes 52
3.2.4 Combined PStat3 and Klf4 bypass Nanog in reprogramming 55
3.3 Discussion 59
3.3.1 Modulation of LIF signal transduction by Nanog 59
3.3.2 Multifaceted relationship between Nanog and the environment 60
3.3.3 Klf4-directed synergy between Nanog and PStat3 62
3.3.4 Role and conservation of the novel *Klf4* enhancer RNA 63
CHAPTER 4: GENERATION OF A TRACTABLE IDENTITY TRANSITION SYSTEM 67
4.1 Introduction 67
4.2 Results 69
4.2.1 Interaction between TFs and signalling environment during reprogramming 69
4.2.2 Isolation of productive reprogramming intermediates 73
4.2.3 Optimisation of Stat3 activation by GCSF/GY118F 76
4.2.4 Further technical improvements to EpiSC culture and reprogramming 83
4.2.5 Model set of single reprogramming drivers 85
4.3 Discussion 88
CHAPTER 5: TRANSCRIPTIONAL TRAJECTORIES ARE DRIVER-SPECIFIC 89
5.1 Introduction 89
5.2 Results 92
5.2.1 Initiation of naïve network wiring is driver-specific 92
5.2.2 Signal contribution to reprogramming initiation is modulated by the driver 94
5.2.3 Single-cell RNAseq defines distinct productive trajectories 96
5.3 Discussion 103
CHAPTER 6: ROUTES ARE MECHANISTICALLY DISTINCT 105
6.1 Introduction 105
6.2 Results 106
6.2.1 Downstream genetic mediators of reprogramming drivers 106

6.2.2 Exogenous signal requirements for each reprogramming route 108
6.2.3 Differential modulation of exogenous signal transduction to target genes 109
6.2.4 Identification of endogenous BMP signalling as a route-specific mechanism 110
6.2.5 BMP signalling is required for naïve pluripotency establishment *in vivo* and during human resetting 113
6.3 Discussion 117
6.3.1 Mechanistic distinctions between reprogramming routes 117
6.3.2 BMP signalling in naïve pluripotency establishment 118
CHAPTER 7: ROUTE RECONCILIATION WITH A UNIFYING TRANSITION LOGIC 121
7.1 Introduction 121
7.2 Results 122
7.2.1 Defined Oct4 level is a common feature of all routes 122
7.2.2 Oct4 maintenance is an active feature of productive reprogramming 124
7.2.3 Fixed Oct4 expression is sufficient for naïve instruction in minimal conditions 126
7.2.4 Reconciliation of route differences with common Oct4 maintenance 128
7.2.5 Fine-tuned Oct4 expression is a pivotal feature for somatic reprogramming 132
7.3 Discussion 135
7.3.1 Unification of reprogramming logics 135
7.3.2 Applicability to somatic cell reprogramming 136
7.3.3 Resolving the iKlf2 paradox 137
CHAPTER 8: GENERAL DISCUSSION 139
8.1 General conclusions and implications 139
8.1.1 Principles of cell identity transitions 139
8.1.2 Interplay between transcription factors and environmental signals 141
8.1.3 Network differences for installation vs maintenance of naïve pluripotency 143
8.2 Suggestions for future work 144
8.2.1 Genomic binding and role of Oct4 during identity transitions 144
8.2.2 Balancing of the naïve pluripotency network in ESCs 145
8.2.3 Does Klf2 play a post-implantation role in mesodermal lineage specification? 146
8.2.4 Sox2 mRNA regulation and Oct/Sox partner swapping 147
8.2.5 Transition factors during regenerative fate transitions *in vivo*? 149
CHAPTER 9: BIBLIOGRAPHY 151
APPENDIX 171

List of Figures

Figure 1.1.1: Early murine development....6
Figure 1.2.1: Signals governing naïve pluripotency....9
Figure 1.2.2: Naïve pluripotency TF network and 2iLIF signal inputs....14
Figure 1.2.3: Primed and naïve pluripotency....20
Figure 1.2.4: Stat3 activation by endogenous LIFR vs by transgenic GY118F....25
Figure 3.1.1: Nanog and PStat3 hold a special status amongst naïve pluripotency TFs....42
Figure 3.2.1: Nanog amplifies Stat3 phosphorylation and negatively correlates with Socs3 44
Figure 3.2.2: Validation of dox-inducible Nanog transgene....45
Figure 3.2.3: Nanog binds the Socs3 promoter and represses its transcription....46
Figure 3.2.4: Nanog modulates the response of Klf4 to LIF....47
Figure 3.2.5: Identification of novel enhancer RNA at the *Klf4* locus....49
Figure 3.2.6: *K4eRNA* responds to Nanog and PStat3 in the same manner as Klf4....50
Figure 3.2.7: Nanog and PStat3 synergistically upregulate *K4eRNA* and Klf4....51
Figure 3.2.8: Generation of *Nanog-/-* embryo-derived EpiSCs....53
Figure 3.2.9: Nanog and PStat3 induce rapid reactivation of naïve genes in EpiSCs....55
Figure 3.2.10: Combined PStat3 and Klf4 bypass Nanog in reprogramming....57
Figure 3.2.11: Acquisition of naïve gene expression signature in *Nanog-/-* iPSCs....58
Figure 3.3.1: Newfound role of Nanog in the modulation of LIF/Stat3 signal transduction ...60
Figure 3.3.2: Condition-dependent relationship between Nanog and epigenetic regulation?61
Figure 3.3.3: Synergistic response of Klfs to Nanog and PStat3 in human ESCs....63
Figure 3.3.4: Conservation of the *K4eRNA* locus....65
Figure 4.1.1: Productive intermediates must be isolated from the bulk....68
Figure 4.2.1: Reprogramming gene expression profiles in FA or 2iLIF....70
Figure 4.2.2: Network modulation by timing of signal switch in reprogramming....72
Figure 4.2.3: Validation of *Rex1::*dGFP reporter for isolation of productive intermediates....75
Figure 4.2.4: Lessons from the naïve state regarding appropriate Stat3 activation level....78
Figure 4.2.5: Balance between Stat3 and GY118F in EpiSC reprogramming....80
Figure 4.2.6: High-efficiency low-GY118F reprogramming....82
Figure 4.2.7: Improved EpiSC reprogramming in hypoxia....84
Figure 4.2.8: Model set of single reprogramming drivers....86
Figure 4.2.9: Functional validation of single-driver *Rex1::*dGFP EpiSC reprogramming....87
Figure 5.1.1: Are routes distinct or equivalent within the productive subpopulations?....90
Figure 5.1.2: Overview of reprogramming phases for *Rex1*::dGFP EpiSCs....91
Figure 5.2.1: Initiation of naïve network wiring is driver-specific....93
Figure 5.2.2: Signal contribution to reprogramming initiation is modulated by the driver....95

Figure 5.2.3: Single-cell RNAseq of productive reprogramming intermediates 97
Figure 5.2.4: Kinetics of naïve network induction in productive single cells 98
Figure 5.2.5: Expression of mesodermal markers in single iKlf2 intermediates 99
Figure 5.2.6: Developmental context of productive reprogramming intermediates 101
Figure 5.2.7: Nanog+Gata6+ iPStat3 intermediates emulate developmental progression . 102
Figure 5.3.1: Summary of transcriptional trajectories for reprogramming routes 103
Figure 6.2.1: Route-specific dependence on Klf2 as a reprogramming mediator 107
Figure 6.2.2: Route-specific requirements for exogenous signals..... 108
Figure 6.2.3: Synergy with Esrrb amplifies LIF/Stat3 target gene upregulation 109
Figure 6.2.4: BMP signalling requirements for naïve pluripotency specification 111
Figure 6.2.5: Endogenous BMP signalling is dispensable for resultant iPSCs..... 112
Figure 6.2.6: Effect of DMH2 treatment on resetting of primed human ESCs..... 114
Figure 6.2.7: Timed addition of DMH2 to embryos disrupts pluripotency induction 116
Figure 6.3.1: Summary of mechanistic requirements for each reprogramming route 117
Figure 7.2.1: Oct4 maintenance is a unifying feature of productive reprogramming..... 122
Figure 7.2.2: Oct4 maintenance during reprogramming is not to be taken for granted 123
Figure 7.2.3: Oct4 maintenance correlates with and is required for reprogramming..... 125
Figure 7.2.4: Fixed Oct4 expression is sufficient for naïve instruction 127
Figure 7.2.5: Immediate impact of each driver on Oct4 expression 129
Figure 7.2.6: Reconciliation of naïve induction logic between different routes..... 131
Figure 7.2.7: Conservation of Oct4 vs Sox2 regulation between species 131
Figure 7.2.8: Derivation of FixedOct4 somatic cells by differentiation in chimeras 132
Figure 7.2.9: Fine-tuned Oct4 expression is a pivotal feature for somatic reprogramming . 134
Figure 7.3.1: Convergence on precise Oct4 expression to access naïve pluripotency..... 136
Figure 8.2.1: Pou/Sox family responses to Klf2 induction in EpiSCs 148

List of Tables

Table 1: Applied Biosystems TaqMan RT-qPCR assays 36
Table 2: Applied Biosystems TaqMan RT-qPCR custom assays..... 36
Table 3: ChIP primers..... 37

List of Common Abbreviations

2i	dual inhibition of Mek/Erk by PD0325901 and GSK3 by CHIR99021
2iLIF	2i together with LIF in N2B27 medium
Aza	5-Azacytidine
BMP	bone morphogenetic protein
Bsd	blasticidin
CH	CHIR99021
ChIP	chromatin immunoprecipitation
dGFP	destabilised green fluorescent protein with a half-life of 2 hours
Dox	doxycycline
Dnmt	DNA methyltransferase
E	embryonic day
ECC	embryonal carcinoma cell
EpiSC	post-implantation epiblast stem cell
Erk	extracellular signal regulated kinase
ESC	embryonic stem cell
EV	empty vector
F	floxed allele
FA	Fgf2 together with ActivinA in N2B27 medium
FCS	foetal calf serum
Fgf	fibroblast growth factor
GCSF	granulocyte colony stimulating factor
GRN	gene regulatory network
GSK3	glycogen synthase kinase 3
GY118F	chimeric GCSF/GP130 receptor construct with point mutation
H3K4me3	histone H3 trimethylated on lysine 4
H3K27me3	histone H3 trimethylated on lysine 27
hESC	human embryonic stem cell
ICM	inner cell mass
iGOI	inducible gene-of-interest
iPSC	induced pluripotent stem cell
iPStat3	cells constitutively expressing GY118F
Jak	Janus activated kinase
K2N	Klf2 and Nanog transgenes
K4eRNA	RNA expressed from the *Klf4* enhancer
KSR	knockout serum replacement

KD	knock-down
LIF	leukaemia inhibitory factor
Lifr	LIF receptor
LOESS	smooth curve local regression
mC	5-methylcytosine (DNA base)
Mek	mitogen activated protein kinase
N2B27	basal serum-free medium
OE	constitutive overexpression
PB	piggyBac vector
PCA	principle component analysis
PD	PD0325901
PGC	primordial germ cell
PI3K	phosphoinositide 3-kinase
PrE	primitive endoderm
pre-iPSC	partially reprogrammed intermediate of somatic cell reprogramming
PStat3	Stat3 phosphorylated on tyrosine-705
RT-qPCR	real-time quantitative PCR
sc	single-cell
SD	standard deviation
SH2	Src homology-2 domain
siRNA	short interfering RNA
Socs	suppressor of cytokine signalling
Stat	signal transducer and activator of transcription
WT	wild-type

CHAPTER 1: GENERAL INTRODUCTION

1.1 Cellular identity

1.1.1 What is cellular identity?

Multicellular animals are built from an astounding array of cell types. Each cell has properties appropriate to its particular role as a building block and functional element of the whole animal. For example, sensory neurons must contain appropriate receptors and apparatus to transduce stimuli into action potentials along axons, cardiomyocytes require contractility and fatigue-resistance, while hepatocytes contain abundant endoplasmic reticuli to support protein synthesis and export. Not only does each of the >200 cell types have to be specified during development and tissue homeostasis, but they must also work together to produce and support the coherent organism.

Gradual loss of unneeded genes during development would provide a simple explanation for fidelity in fate restriction and inheritance of identity memory from mother to daughter cell. Whilst programmed genome rearrangement occurs to a limited extent in a handful of species (Smith, 2017b), such a paradigm is incompatible with the remarkable regulative responses following injuries, developmental perturbations, or during metamorphosis. A paradigm shift in our comprehension of cell fate determination was provided by the landmark work of John Gurdon, in which he demonstrated that terminally differentiated adult cells in fact possess all the genetic information necessary to make an entire animal (Gurdon, 1962). Thus, since different cells contain the same information, the key is how they decide and remember which subset of the genome to use.

1.1.2 What determines cellular identity?

We thus move up a level, to consider genomic regulation as the governor of cellular identity. Since the genome is the same in almost every cell, there must be other forms of informational input that instruct cellular diversification. In single-celled organisms, we know that expression of a given gene depends on the presence and activity of transcriptional machinery, and that this can be modulated in response to environmental cues (Lee et al., 2002; Reznikoff, 1989). Extrapolating to multicellular animals, a brute-force solution could be that, for each cell type, there exists a particular signal responsible for turning on a unique master transcriptional

regulator, which in turn activates an identity-specific gene-set. Genomic sequencing in the 1990s exposed a fundamental flaw in this simplistic hypothesis: there simply were not enough genes in the genome to permit such linear identity computation in addition to supportive biochemical functions (Lander et al., 2001). Correspondingly, forced expression of single transcription factors (TFs) affects identity in a context-dependent manner, usually altering phenotypes only to those of closely related cell types (Farah et al., 2000; Graf and Enver, 2009; Schäfer et al., 1990).

An immense scientific effort has led us to today's appreciation of the elegant efficiency with which differential use of the same genome generates the spectacular diversity of cellular form and function. The finite number of TFs and signalling pathways are used and re-used in different combinations, concentrations, spatiotemporal and epigenetic contexts (Section 1.1.3). These parameters provide the additional layers of informational complexity required to generate distinct identities: context and history modulate how cells compute inputs and execute decisions within the framework of the single genome. Whilst we now grasp this in a general sense, our understanding of the specifics is incipient and constitutes one of the most exciting areas of research in this post-genomic era.

1.1.3 Identity stability

Once a cellular identity is established, how is it stably maintained over time and, if appropriate, divisions? How does a cell remember its identity without being continually exposed to active specification events? Identity fidelity is critical to healthy organismal function, with misregulation leading to developmental defects and diseases such as cancer.

Gene regulatory networks (GRN) specify the set of interacting genes and proteins responsible for identity governance, many of which are TFs. GRN architecture dictates the logic by which these variables interact, and can explain many aspects of gene expression regulation across space and time. Networks of cross-regulating TFs can self-reinforce a stable identity state while repressing others (Davidson et al., 2002; Huang et al., 2007; Levine and Davidson, 2005; Swiers et al., 2006). Network properties such as hysteresis can also protect identity by rendering established states insensitive to noise, and facilitate the paradigm of signal reuse during developmental patterning. A network state dependent on historical as well as ongoing inputs contains a 'memory', allowing the same perturbatory signal to fulfil a different function at a later timepoint without disrupting past decisions (Briscoe and Small, 2015; Herrera-Delgado et al., 2018).

A cellular identity with a stable GRN can be thought of as an attractor, occupying a local minimum in an 'energetic' landscape of gene expression space (Kauffman, 1993; Enver et al., 2009). Modelling of cellular identity in dynamical systems terms formalises qualitative metaphors of stable cell states as 'valleys' separated by 'hills' in Waddington's landscape (Huang, 2012; Jaeger and Monk, 2014; Moris et al., 2016; Waddington, 1957). Conceptualisation of cell identity as a solution (attractor) to a dynamical model explains the robust and discrete nature of cellular identity: when subjected to noise or minor fluctuations, the system will tend to gravitate back to that equilibrium state.

Epigenetic regulation is a complementary approach by which cells restrict genome usage to the required identity subset and protect the identity from noise (Nashun et al., 2015; Pujadas and Feinberg, 2012). By modifications to the DNA and the chromatin packaging it, the accessibility of gene loci and their subnuclear localisation can be flexibly regulated (Goldberg et al., 2007; Ho et al., 2014; Nashun et al., 2015; Xie et al., 2013). This reinforces cell identity decisions: most TFs cannot engage effectively with heterochromatin, and thus the programs for unwanted identities can be robustly silenced (Voss and Hager, 2014). Since epigenetic modifications are stably inherited during subsequent cell divisions, this provides a memory long after the specifying event has occurred, and can also restrict the range of available fates in any subsequent decisions (Jost, 2014; Shipony et al., 2014; Zhu et al., 2013). The epigenetic landscape is thus a crucial parameter for the installation and stable maintenance of different identities from a single genome, modulating the context in which TFs operate.

1.1.4 Cell identity transitions

Cell identity stability must be balanced with the means to reliably change identity when appropriate. During animal development, a series of identity transitions occurs with exquisite precision, producing a functional organism from a single fertilised egg. Not only must the required cell types be generated, but also choreographed in the appropriate space, time and quantities. Identity transitions must also be instigated on demand, in response to events and the environment during tissue homeostasis and repair.

TFs are understood to be key regulators of cellular identity (Section 1.1.2). Impressively, ectopic imposition of one or a few TFs can suffice to coordinate a fate change. For example: MyoD drives the conversion of murine fibroblasts to myocytes (Davis et al., 1987); Antennapedia misexpression in Drosophila elicits the formation of legs instead of antennae (Schneuwly et al., 1987); delivery of Oct4, Sox2, Klf4 and cMyc reprograms somatic cells to

pluripotent stem cells in a variety of vertebrate species (Friedrich Ben-Nun et al., 2011; Ogorevc et al., 2016; Takahashi and Yamanaka, 2006). This remarkable ability of a handful of TFs to orchestrate the global changes required to change cell type is a testament to their causative role in identity transition, and demonstrates that in certain cases activation of one or a few nodes can trigger installation of a complete identity TF network.

What, then, regulates the identity-specifying TFs themselves? In addition to cross-regulation within networks, the expression and activity of the TFs themselves can be induced or repressed by external signals. As introduced in Section 1.1.2, permutations and concentrations of signals/TFs regulate cell identity specification and maintenance. However, for a signal to play an instructive role in an identity transition, the cell must be competent to transduce it, i.e. the relevant machinery must be expressed, which is itself a function of the transcriptional state of the cell. Assuming transducability, signal interpretation is in turn dependent on the underlying TF network architecture (Briscoe and Small, 2015). Signalling and transcriptional parameters thus engage in a complex interplay in order to drive and direct identity transitions.

Forced fate changes *in vitro* are usually inefficient and heterogeneous, indicative of barriers between identities even when appropriate TFs and signals are supplied. Other cell-intrinsic properties can determine whether and how a signal or TF is interpreted, such as the epigenetic accessibility of target genes for perturbation. Interaction between TFs and the epigenetic landscape is bidirectional, with the chromatin structure modulating TF engagement yet in turn being altered by TF-recruited modifiers (Nashun et al., 2015). Thus, not only must the driving factors for an identity change coordinate remodelling of the epigenetic landscape from the source to destination signature, but this in itself actively interacts with the ongoing transition. Epigenetic barriers are regularly encountered during attempts to force fate transition (Chen et al., 2013; Gaspar-Maia et al., 2013; Mikkelsen et al., 2008; Pasque et al., 2012; Sridharan et al., 2013). A privileged class of TFs termed ‘pioneers’ are able to bind closed heterochromatin and thereby initiate locus activation, including POU, FOXA and GATA TF family members (King and Klose, 2017; Zaret and Carroll, 2011). Pioneer factors provide a necessary balance to the nucleosome-rich eukaryotic chromatin structure, and are critical for the initiation of developmental lineages with different expression programs (Kornberg and Lorch, 1999; Spitz and Furlong, 2012; Voss and Hager, 2014).

Cell identity transitions can be formalised according to the attractor model introduced in Section 1.1.3. In these dynamical systems terms, a change in identity entails the transition from one attractor state to another, overcoming an energetic barrier between states. Whether

or not a cell succeeds in such a transition can be described by probability functions, whereby reduction of the barrier renders the transition more likely. Whilst attractor states can withstand some degree of noise, strong perturbations can destabilise the source attractor and modulate the relative accessibilities of alternative solutions in gene expression space, thus triggering a cell identity transition (Ferrell, 2012; Huang et al., 2007; Kauffman, 1993). In this way, cues can provide cells with both the driving force and the directional impetus in order to execute appropriate identity transitions.

1.1.5 Identity potency

The potency of a cell refers to the compilation of cell types that it can generate in response to appropriate cues, either directly or indirectly via progenitor stages. High potency states exist transiently during early development, whilst multi/bi/unipotent stem cells play critical roles in tissue growth, homeostasis and repair (Simons and Clevers, 2011). Cell types with a degree of potency are particularly intriguing: not only do such cells have an identity in themselves with the properties they need to fulfil their functions, but a major aspect of their function itself is to retain access to other fates.

At the beginning of mammalian development, the fertilised zygote is considered totipotent because it ultimately gives rise to all cell types of the embryo and the supporting extraembryonic tissues such as placenta and yolk sac. However, we would not formally consider the zygote to be a stem cell, because it does not self-renew: its very nature necessitates a transient existence. Following a period of proliferation and specification of the extraembryonic trophectoderm and primitive endoderm lineages, pluripotent cells constitute the epiblast of the blastocyst (Gardner et al. 1971, 1979, 1985) (Figure 1.1.1A). Pluripotency is defined as the potential to form all cell types of the embryo-proper and by extension the adult, including the germ lineage from which gametes originate.

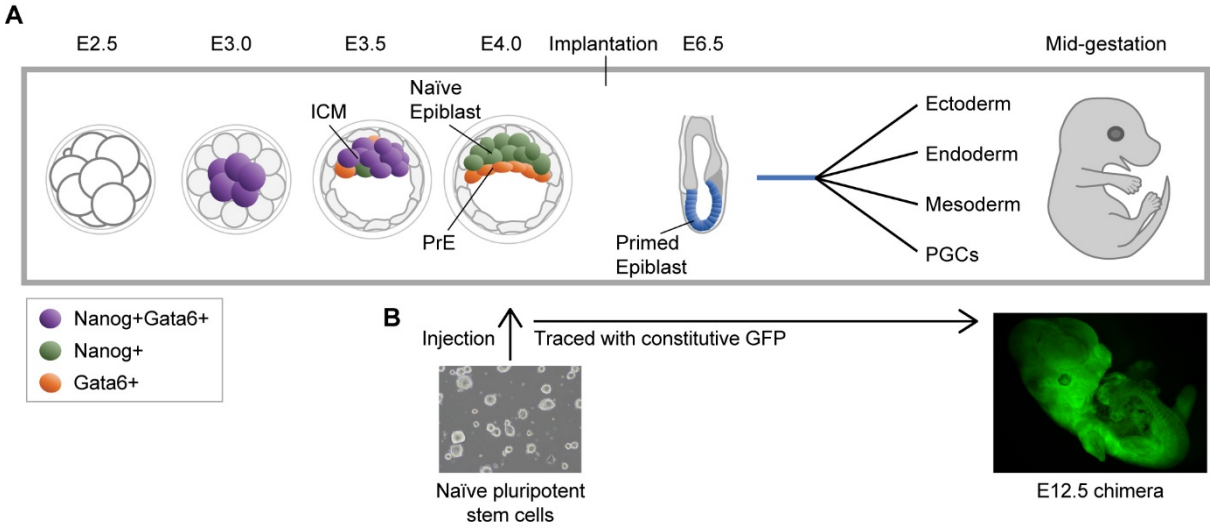


Figure 1.1.1: Early murine development

(A) Early mammalian development has been best characterised in the mouse model. Knowledge accumulated from murine systems over the last 50 years has led to our current appreciation that pluripotency exists *in vivo* as a continuum of states from embryonic day (E)3.5–8.0 (Morgani et al., 2017; Smith, 2017a), during which time the epiblast forms and progresses. Following specification of the trophectoderm externally (Sutherland et al., 1990), the inner cell mass (ICM) at E3.25 co-expresses both Nanog and Gata6 TFs. By E3.75 Nanog or Gata6 are expressed in a mutually exclusive 'salt and pepper' pattern (Chazaud et al., 2006; Plusa et al., 2008). The emergent Nanog+ epiblast and Gata6+ primitive endoderm (PrE) lineages undergo sorting and consolidation, such that by the E4.0 pre-implantation blastocyst they are segregated both physically and in terms of fate. In peri- and post-implantation stages, epiblast progression through formative then primed pluripotency capacitates it for subsequent differentiation, triggered by signals such as fibroblast growth factor (FGF), bone morphogenetic protein (BMP), Wnt and Nodal.
(B) When injected into the pre-implantation blastocyst, naïve pluripotent stem cells contribute to all embryonic lineages, including to primordial germ cells (PGCs).

1.2 Pluripotency

1.2.1 Naïve pluripotency capture and culture

The naïve pluripotent identity captures the imagination due to both its conceptual elegance and the promise it holds as a tool for biological research and therapeutic applications. Since naïve pluripotent cells possess an unbiased potential to make all lineages of the developed organism, propagation of this identity *in vitro* offers a putative source material for any required tissue via differentiation, as well as a window into understanding the molecular basis of this remarkable cellular state itself.

Pluripotency was first captured *in vitro* in the 1970s as murine embryonal carcinoma cells (ECCs) from testicular and embryo-derived carcinomas. Whilst a subset of ECCs were capable of forming multiple healthy tissues, their heterogeneous and tumorous origin was problematic (Kahan and Ephrussi, 1970; Martin and Evans, 1975). By applying optimised ECC culture conditions (batch-tested foetal calf serum (FCS) and co-culture with mitotically inactivated fibroblastic feeders), pluripotent cell lines were successfully derived directly from murine blastocysts (Evans and Kaufman, 1981; Martin, 1981). These embryonic stem cells (ESCs) simultaneously retain the capacity for indefinite self-renewal *in vitro* and the potential to resume normal development. Following blastocyst injection, ESCs contribute to all tissues of chimeric adults, including the germ lineage (Bradley et al., 1984) (Figure 1.1.1B). Remarkably, this permits transmission of the ESC genome to the next generation, revolutionizing the engineering of mouse genetic models (Robertson et al., 1986; Thomas et al., 1986; Thompson et al., 1989).

Deconvolution of the signal milieu provided by FCS and feeders led to recognition of two key ligands for ESC culture: the cytokine leukaemia inhibitory factor (LIF) from feeders (Smith et al., 1988; Williams et al., 1988), and bone morphogenetic protein 4 (BMP4) from FCS (Ying et al., 2003). Despite promotion of mesendodermal differentiation by BMP4 in other contexts, when applied together LIF and BMP promote ESC maintenance by blocking differentiation (Wiles and Johansson, 1999; Ying et al., 2003). In the presence of LIF, feeders are no longer required and from hereon all cultures are assumed to be feeder-free unless explicitly stated otherwise.

LIF signals through a bipartite receptor comprised of LIF receptor (Lifr) and glycoprotein 130 (Gp130), triggering activation of Janus kinase-2 (Jak2) associated with the intracellular domain of Gp130. Three signalling pathways lie downstream of Jak2 in ESCs: phosphatidyl

inositol-3 kinase (PI3K), extracellular signal-related kinase (Erk), and signal transducer and activator of transcription 3 (Stat3). Curiously, Erk signalling is actually antagonistic to self-renewal (Burdon et al., 1999) whereas Stat3 is responsible for the beneficial effect of LIF on ESC maintenance (Boeuf et al., 1997; Matsuda et al., 1999; Niwa et al., 1998). In order to activate Stat3, Jak2 phosphorylates the intracellular domain of GP130, causing recruitment of latent cytoplasmic Stat3 to Gp130 via the Stat3 Src homology-2 (SH2) phosphotyrosine-binding domain. Jak2 then phosphorylates tyrosine-705 of Stat3 (PStat3), allowing SH2-mediated homodimerization of PStat3 and consequent nuclear translocation and activity (Johnston and Grandis, 2011; Kisseleva et al., 2002). Suppressor of cytokine signalling 3 (Socs3) binds to Gp130-Jak2 complexes, prevents Stat3 binding to Jak2, and thus inhibits Stat3 phosphorylation and activation (Kershaw et al., 2013). Socs3 transcription is rapidly and strongly induced by PStat3, to form a classic negative feedback loop in which PStat3 and Socs3 levels oscillate in antiphase (Yoshiura et al., 2007) (Figure 1.2.4).

ESCs cultured in FCS+LIF or BMP4+LIF exhibit heterogeneity both morphologically and in terms of marker expression, and spontaneous differentiation is not uncommon (Chambers et al., 2007; Festuccia et al., 2012; Hayashi et al., 2008; Reynolds et al., 2012; Smith, 2001; Toyooka et al., 2008). Since ESC differentiation involves Fgf-stimulation of Erk mitogen-activated protein kinases (Kunath et al., 2007), inhibitors were tested to block autocrine Fgf signalling. Small molecule inhibition of Fgf receptors or of Erk-activating Mek1/2 permits ESC self-renewal in the presence of LIF only (Ying et al., 2008). Culture robustness is further improved by concomitant inhibition of glycogen synthase kinase 3 (Gsk3). Strikingly, Mek/Erk inhibition by PD0325901 (PD), Gsk3 inhibition by CHIR99021 (CH) and exogenous LIF ligand applied in any pairwise combination is sufficient to support ESC self-renewal (Dunn et al., 2014; Ying et al., 2008). Since culture with the two inhibitors (2i) alleviates the requirement for Stat3-agonist LIF, this finally permitted the generation of *Stat3-/-* ESCs and derivation of ESCs from non-permissive mouse strains in which the Erk:Stat3 activation ratio downstream of LIF is too high (Ohtsuka and Niwa, 2015; Ying et al., 2008). However, addition of LIF to wild-type ESCs increases clonogenicity, confers robust and rapid expansion, and enhances ESC-appropriate oxidative metabolism (Carbognin et al., 2016; Dunn et al., 2014; Wray et al., 2010).

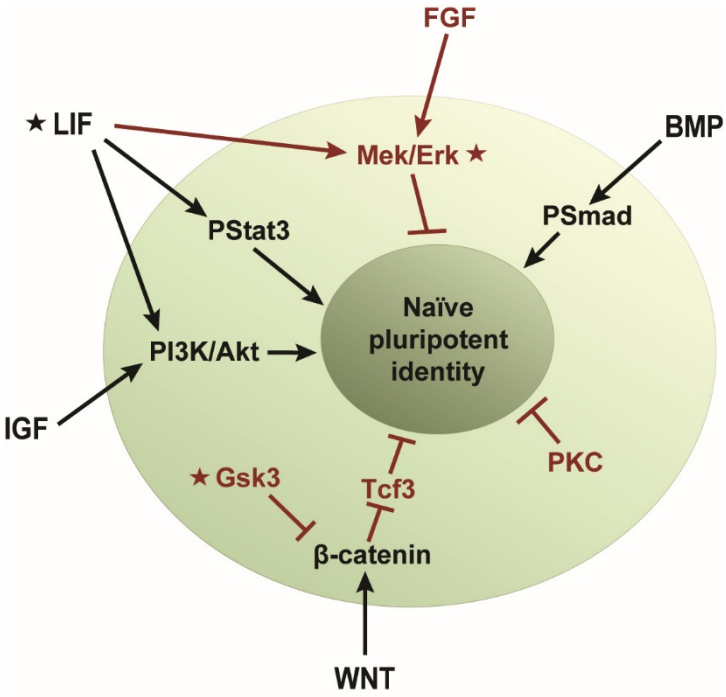


Figure 1.2.1: Signals governing naïve pluripotency

LIF promotes naïve pluripotency self-renewal via activation of Stat3 and PI3K/Akt signalling (Boeuf et al., 1997; Matsuda et al., 1999; Niwa et al., 1998). PI3K/Akt is also activated by insulin-like growth factors (IGFs). Mek/Erk signalling lies downstream of both LIF and FGF ligands, and drives differentiation (Burdon et al., 1999). In serum-containing medium, BMP signalling inhibits differentiation via Smad activity (Ying et al., 2003). PKC inhibitors promote ESC self-renewal (Dutta et al. 2011) but are not routinely applied. β-catenin activity inhibits Tcf3, thus alleviating pluripotency gene repression (Lyashenko et al., 2011; Wray et al., 2011). The 'ground-state' of naïve pluripotency is captured in 2iLIF medium, the components of which are starred: LIF ligand; Chiron to inhibit Gsk3; PD03 to inhibit Mek/Erk signalling.

2i together with LIF (2iLIF) yields an optimal culture of naïve ESCs (starred in Figure 1.2.1). Not only does this regime have benefits in terms of molecular definition, but importantly the homogeneity and undifferentiated status of resultant ESCs at clonal density is markedly improved over conventional FCS+LIF conditions (Wray et al., 2010, 2011). Unlike in pairwise combinations, provision of all three 2iLIF components fully suppresses lineage marker expression and differentiation biases (Hackett et al., 2017). Together this demonstrates that thorough insulation from pro-differentiation signals is sufficient for robust ESC maintenance. On this basis, a prevailing hypothesis is that ESC self-renewal is a cell-autonomous program and that 2iLIF-cultured ESCs embody the 'ground-state' of unbiased naïve pluripotency (Silva and Smith, 2008; Wray et al., 2010; Ying et al., 2008).

Since naïve pluripotency arises only transiently during development, the indefinite self-renewal capacity of ESCs raises questions regarding their *in vivo* correspondence. The ability of blastocyst-injected ESCs to seamlessly re-enter full development indicates functional

equivalence to the pre-implantation epiblast, the founder population of the embryo (Gardner et al. 1971, 1979, 1985) (Figure 1.1.1B). Whilst ESCs can be derived in 2iLIF from the ICM of mid- to mature-blastocyst stages, the resultant cell lines have functional, transcriptional, signalling and epigenetic signatures which match that of the naïve epiblast just prior to implantation (Boroviak et al., 2014; Nichols et al., 2009). Although it cannot formally be excluded that other changes occur during ESC derivation and adaptation, they do recapitulate the defining features of the naïve epiblast and are thus considered its *in vitro* counterpart. A parallel can be drawn to the natural long-term maintenance of naïve pluripotency during LIF/Stat3-dependent facultative diapause of murine embryos; perhaps this innate ability underpins the relative ease of murine naïve ESC derivation compared to other species (Nichols and Smith, 2009; Nichols et al., 2001).

1.2.2 The transcription factor network governing naïve pluripotency

The TF network governing murine naïve pluripotency is arguably one of the best defined of all mammalian cell types. This is largely thanks to the ability of ESCs to undergo indefinite clonogenic self-renewal, facilitating genetic manipulation and providing unlimited material for analysis, coupled with stringent functional assays such as clonal assays and chimeric contribution (Martello and Smith, 2014; Tam and Rossant, 2003). Over the last 30 years, many pluripotency regulators have been identified and interactions between them characterised.

Oct4, Sox2 and Nanog are thought to comprise the core of the pluripotency TF network due to their expression in both naïve and primed pluripotent states, and each is essential for pluripotency establishment during embryonic development. Oct4 was the first pluripotency factor to be identified, based on its expression pattern in the early embryo and germ cells (Okamoto et al., 1990; Schöler et al., 1990; Yeom et al., 1996). Oct4 is required for establishment of naïve pluripotency *in vivo*; without it, the ICM is allocated normally but then generates trophectoderm instead (Nichols et al., 1998). As a member of the POU family, Oct4 contains two DNA-binding domains (POU and homeodomain) connected by a unique linker proposed to recruit epigenetic modulators to Oct4 target genes (Esch et al., 2013), and is a rare example of a ‘pioneer’ TF able to bind its targets even when their chromatin is closed (King and Klose, 2017; You et al., 2011).

Curiously, either deletion or overexpression of Oct4 *in vitro* leads to ESC differentiation (Niwa et al., 2000; Shimozaki et al., 2003; Thomson et al., 2011; Zeineddine et al., 2006). Oct4 expression at wild-type-level permits either ESC self-renewal or multilineage differentiation

depending on the signalling environment, whereas a low level of Oct4 locks ESCs in self-renewal and prevents differentiation *in vitro* or *in vivo* (Karwacki-Neisius et al., 2013; Radzisheuskaya et al., 2013). Thus, while Oct4 is essential for naïve pluripotency establishment and maintenance, its action is highly dose-sensitive and context-dependent. Consistent with its requirement for appropriate differentiation, Oct4 is expressed in multiple lineages of the early embryo up to the late somite stage (DeVeale et al., 2013; Downs, 2008; Osorno et al., 2012). Considering this remarkable set of phenotypes, we understand surprisingly little about the molecular mechanisms of Oct4 action, perhaps due to the inherent difficulty of perturbing it independently of the identity context.

Sox2 is a SRY-box TF that, like Oct4, is essential for pluripotency establishment *in vivo* (Avilion et al., 2003) or ESC maintenance *in vitro* (Masui et al., 2007). Sox2 and Oct4 proteins physically interact (van den Berg et al., 2010; Gagliardi et al., 2013; Pardo et al., 2010) and cooperatively regulate shared target genes (Ambrosetti et al., 1997; Chen et al., 2008; Kuroda et al., 2005; Nakatake et al., 2006; Nishimoto et al., 1999; Tokuzawa et al., 2003; Tomioka et al., 2002; Yuan et al., 1995). However, constitutive Oct4 expression can rescue Sox2 null ESCs, indicating that Oct4 can maintain pluripotency Sox2-independently and suggesting that a key role of Sox2 is to support Oct4 expression (Masui 2007). Like Oct4, Sox2 action is dose-dependent, with overexpression driving differentiation (Kopp et al., 2008; Zhao et al., 2004). However the Sox2 expression profile differs from that of Oct4: in addition to expression in the early embryo (Avilion et al., 2003; Keramari et al., 2010), it is also expressed in a variety of later embryonic and adult lineages including neurectoderm, gut endoderm and epithelial tissues (Sarkar and Hochedlinger, 2013).

Recently, a ‘seesaw’ model of pluripotency regulation was proposed, in which opposing pro-mesendoderm and pro-neurectoderm activities of Oct4 and Sox2 respectively balance to a metastable pluripotency state (Loh and Lim, 2011; Shu et al., 2013; Teo et al., 2011; Thomson et al., 2011; Wang et al., 2012). However, this model is difficult to reconcile with the lack of lineage gene expression in 2iLIF (Hackett et al., 2017), the rescue of *Sox2-/-* ESCs by constitutive Oct4 (Masui et al., 2007), the robust self-renewal of *Sox2+/+* low-Oct4 ESCs (Radzisheuskaya et al., 2013), the multilineage differentiation triggered by Oct4 or Sox2 overexpression in various conditions (Kopp et al., 2008; Niwa et al., 2000; Shimozaki et al., 2003; Thomson et al., 2011), or the expression and roles of Oct4 and Sox2 in multiple lineages during development (Clavel et al., 2012; DeVeale et al., 2013; Downs, 2008; Gontan et al., 2008; Pevny and Nicolis, 2010; Que et al., 2007; Wegner and Stolt, 2005). Again, perhaps, it is a question of context: throughout development, different interaction partners of and differential pairing between Oct/Sox family members profoundly modulate their roles (Adachi

et al., 2013; Aksoy et al., 2013; Buecker et al., 2014; Kondoh and Kamachi, 2010; Yang et al., 2014).

Nanog is a homeodomain-containing TF unique to vertebrates (Chambers et al., 2003). Like Oct4 and Sox2, Nanog is required for naïve pluripotency establishment *in vivo* (Mitsui et al., 2003; Silva et al., 2009) and its expression correlates more closely with *in vivo* pluripotency than that of Oct4 or Sox2 (Osorno et al., 2012; Silva et al., 2009). Curiously, once naïve pluripotency is established Nanog becomes dispensable for its maintenance, allowing the generation of *Nanog-/-* ESCs that can contribute to all somatic tissues upon blastocyst injection, but only in a limited manner to the germline (Chambers et al., 2007; Zhang et al., 2018). However, *Nanog-/-* and Nanog-low ESCs are more prone to differentiation than wild-type ESCs (Chambers et al., 2003, 2007; MacArthur et al., 2012; Mitsui et al., 2003). Unlike Oct4/Sox2, Nanog overexpression is beneficial for pluripotency maintenance, supporting ESC self-renewal in minimal culture conditions such as serum without LIF (Chambers et al., 2003; Mitsui et al., 2003) or even in basal medium (N2B27) only (Ying et al., 2003). Whilst Oct4, Sox2 and Nanog cross-regulate each other and are considered the core of the pluripotency network (Young, 2011), the complex and at times contrasting phenotypes outlined above indicate complementary as well as shared roles. Indeed, a portion of Nanog-bound genes are co-occupied by Oct4/Sox2 in ESCs, but many are not (Loh et al., 2006; Marson et al., 2008; Nishiyama et al., 2013). Given also their common expression, other parameters must modulate the actions of Oct4/Sox2/Nanog between naïve vs primed pluripotency contexts (see Section 1.2.6).

Naïve-specific TFs integrate with the Oct4/Sox2/Nanog core to form the regulatory network governing naïve pluripotency (Figure 1.2.2). Such factors are expressed in naïve but not primed pluripotent cells both *in vivo* and *in vitro*, and include Esrrb, Gbx2, Klf2, Klf4, Klf5, PStat3, Tbx3, Tfcp2l1 and Tcl1. Whilst single depletion of these factors can generally be tolerated (Festuccia et al., 2012; Ivanova et al., 2006; Jiang et al., 2008; Martello et al., 2012; Ye et al., 2013; Ying et al., 2008), ESC self-renewal is most robust when all are present, and combinatorial depletion often triggers collapse of the pluripotent identity (Dunn et al., 2014; Jiang et al., 2008). Like Nanog, individual forced expression of Esrrb, Klf2, Klf4, Tbx3 or Tfcp2l1 can support ESC self-renewal in the absence of LIF or in minimal conditions (Festuccia et al., 2012; Hall et al., 2009; Ivanova et al., 2006; Martello et al., 2012, 2013; Niwa et al., 2009; Ye et al., 2013). Other factors such as Rex1 are expressed specifically in naïve but not primed pluripotent cells, but do not play an active role in naïve identity regulation and instead provide passive identity markers (Kalkan et al., 2017).

There is a great deal of positive regulation between the factors comprising the naïve regulatory network (Chen et al., 2008; Dunn et al., 2014; Marson et al., 2008). In addition to interactions amongst the Oct4/Sox2/Nanog core (Young, 2011), Esrrb has been identified as a direct target of Nanog (Festuccia et al., 2012). Stat3 targets include Klf4, Klf5, Tfcp2l1 and Gbx2 (Bourillot et al., 2009; Hall et al., 2009; Martello et al., 2013; Niwa et al., 2009; Tai and Ying, 2013; Ye et al., 2013). Klf2 is positively regulated by Oct4 (Hall et al., 2009). This intricate set of regulatory interactions was recently formalised in a data-constrained computational model, compatible with known ESC decisions between self-renewal or differentiation (Dunn et al., 2014). Whilst the predictive power of this model was impressive at 70% accuracy, clearly a significant portion of identity computation logic remains to be defined.

Besides instructing the necessary cellular attributes, there are further requirements for naïve pluripotency governance in terms of the properties of the regulatory network itself. Whilst robust ESC self-renewal and tolerance of single factor depletion indicates a stable, flexible and noise-resistant network, this must be balanced with the capacity for timely network dissolution when appropriate for differentiation. The latter requirement has often been overlooked in simplistic models of the pluripotency network as a positive feedback loop (Martello and Smith, 2014; Young, 2011), and surprisingly few repressive interactions have been defined within the pluripotency network. Therefore, whilst it is not a core aim of my thesis, I will endeavour to highlight any putative negative regulatory relationships if appropriate.

1.2.3 Interplay between signal inputs and transcriptional control of naïve pluripotency

Exogenous 2iLIF signals actively interact with this internal TF network in order to sustain the ground state of naïve pluripotency *in vitro*. As outlined in Section 1.2.2, Mek/Erk inhibition by PD assists in naïve pluripotency capture by blocking autonomous Fgf-triggered differentiation. However, PD also positively regulates the naïve network itself: alleviation of Fgf/Erk-mediated suppression leads to upregulation of Klf2, Nanog, Tfcp2l1 and perhaps others (Lanner et al., 2010; Silva et al., 2009; Ye et al., 2013; Yeo et al., 2014), although we do not fully understand the mechanisms.

Gsk3 inhibition by CH leads to increased signalling of several intracellular pathways, including canonical Wnt/βCatenin (Doble and Woodgett, 2003). Since *βCatenin-/-* ESCs cannot be sustained by 2i or CH+LIF, this appears to be the primary pathway by which CH promotes ESC self-renewal (Lyashenko et al., 2011; Wray et al., 2011). Upon Gsk3 inhibition, βCatenin accumulates and binds to nuclear Tcf/Lef TFs (Clevers, 2006). In ESCs, the predominate

factor Tcf3 acts, unusually, as a transcriptional repressor rather than activator (Cavallo et al., 1998; Merrill et al., 2004; Pereira et al., 2006; Wray et al., 2011). Complex formation with βCatenin leads to Tcf3 displacement from DNA (Shy et al., 2013) and consequently to derepression of Tcf3-target genes, which include Esrrb, Nanog, Tfcp2l1 and Klf2 in ESCs (Martello et al., 2012; Ye et al., 2013). Esrrb appears to be the principal functional effector of Gsk3 inhibition in ESCs (Martello et al., 2012), although other factors and/or pathways may well be involved.

As introduced in Section 1.2.2, Stat3 activation mediates the beneficial effect of LIF for ESC maintenance (Boeuf et al., 1997; Matsuda et al., 1999; Niwa et al., 1998). Indeed *Stat3-/-* ESCs exhibit no functional response to LIF (Martello et al., 2013). PStat3 thus forms a bridge between the signalling environment and the TF regulatory network, of which it is a member. Direct positive targets of PStat3 include Klf4, Klf5, Tfcp2l1 and Gbx2 (Section 1.2.2), of which forced Tfcp2l1 expression most effectively phenocopies LIF provision in ESCs (Martello et al., 2013). Interestingly, Tfcp2l1 is a convergent target of all 2iLIF signalling components (Martello et al., 2012, 2013; Ye et al., 2013). In contrast, core pluripotency factors Oct4 and Sox2 are not directly regulated by 2iLIF signals; we thus infer that their expression in naïve or primed pluripotent cells relies on network connectivity with the naïve- or primed-specific factors (Martello and Smith, 2014).

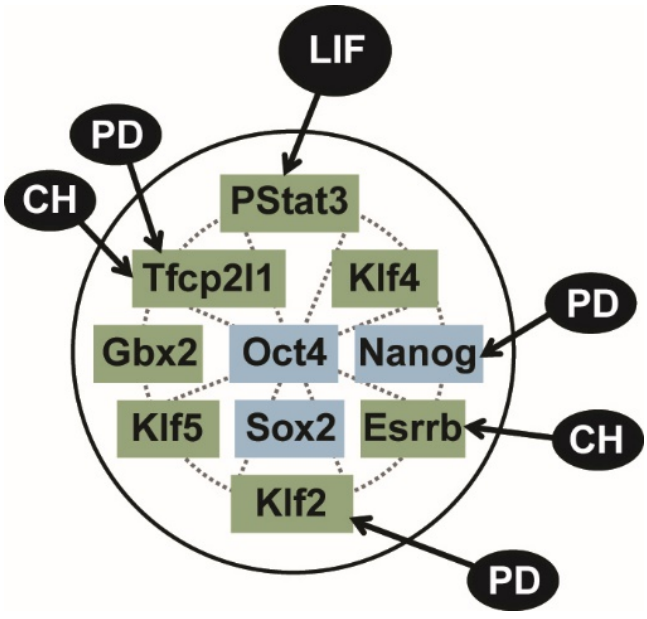


Figure 1.2.2: Naïve pluripotency TF network and 2iLIF signal inputs

Schematic depicting the interconnected TF network governing the naïve pluripotent identity, together with known points of input from 2iLIF signal components. Core and naïve-specific pluripotency factors are indicated in blue and green respectively. PD=PD03; CH=Chiron.

1.2.4 Epigenetic landscape of naïve pluripotency

The naïve epiblast and ESCs cultured in 2iLIF share an unusual epigenetic landscape. With the exception of imprinted and repetitive regions, genome-wide DNA hypomethylation is observed (Ficz et al., 2013; Habibi et al., 2013; Leitch et al., 2013; Smith et al., 2012). These low levels of methylated cytosine in the DNA (30% mC) are attributed to suppression of *de novo* DNA methyltransferases Dnmt3a and Dnmt3b, and to activity of Tet oxidisers of mC. Since mC is generally considered a repressive epigenetic mark, lack thereof is postulated to underpin the unrestricted potential of naïve cells. Likewise, heterochromatin (Ahmed et al., 2010; Efroni et al., 2008; Park et al., 2004) and repressive histone modifications (Hawkins et al., 2010; Loh et al., 2007; Meshorer et al., 2006; Wen et al., 2009) are relatively scarce in the naïve epigenome but accumulate during differentiation.

As with any cell type, naïve pluripotent cells contain an identity-specific set of active enhancers (Buecker et al., 2014; Heintzman et al., 2009; Whyte et al., 2013). However, in addition to this, naïve pluripotent cells exhibit unusually loose genomic packaging overall, rich in open euchromatin (Gaspar-Maia et al., 2009, 2011; Meshorer et al., 2006), activating histone modifications (Krejčí et al., 2009), and with a highly flexible scaffold (Bhattacharya et al., 2009). Together, these epigenetic characteristics may correlate with the extensive and unbiased developmental capacity of naïve pluripotent cells, rendering them free from lineage specification yet allowing access to downstream identity programs on demand.

Another defining epigenetic feature of naïve pluripotency is the presence of two active X chromosomes in female cells (Mak et al., 2004; Okamoto et al., 2004; Rastan and Robertson, 1985), setting the stage for random X chromosome inactivation during subsequent female development. Both the unrestricted epigenetic landscape and X chromosome status are actively instated during naïve pluripotency establishment, and are intimately linked with its TF network strength (Sousa et al., 2018). This provides a strong argument for naïve pluripotency as a gain-of-function during development, rather than passive loss of totipotency from the zygote. Intriguingly, many of these active epigenetic reprogramming events are recapitulated during development of primordial germ cells (PGCs) (Hajkova et al., 2002, 2008; Monk et al., 1987; Seki et al., 2007), coinciding with transient reexpression of pluripotency factors (Seisenberger et al., 2012) and suggesting an intimate link between the pluripotency control network and epigenetic wiping.

ESCs cultured in FCS+LIF exhibit colocalisation of histone H3 trimethylated on lysine 4 (H3K4me3, activating) and histone H3 trimethylated on lysine 27 (H3K27me3, repressive) at

~3000 bivalent promoters (Azuara et al., 2006; Bernstein et al., 2006). Together with heterogeneous TF expression (Section 1.2.2) and promiscuous transcriptional hyperactivity in FCS+LIF (Efroni et al., 2008), this led to the conceptualisation of pluripotency as a metastable state, poised on the verge of lineage specification. However, in 2iLIF this is not the case. When insulated from pro-differentiation signals, transcriptional hyperactivity is not apparent, only ~1000 bivalent genes are observed, and lineage-specific gene expression is repressed (Hackett et al., 2017; Marks et al., 2012). Thus, 2iLIF-cultured ESCs offer an epigenetic 'blank slate', counterparts of the naïve epiblast. Bizarrely, the DNA of ESCs cultured in FCS+LIF is aberrantly (but reversibly) hypermethylated to the level of somatic cells (>70% mC), providing further evidence that their epigenetic status is a culture-induced artefact and urging use of the 2iLIF environment instead (Ficz et al., 2013; Hackett et al., 2017).

1.2.5 Generation of naïve pluripotency by nuclear reprogramming

As introduced in Section 1.1, differentiated cells contain the same genome as ESCs. Therefore, it should theoretically be possible to persuade differentiated cells to alter their usage of the genome to that of the ESC state. This would circumvent the requirement for embryos, and allow the generation of pluripotent cells genetically matched to existing adult organisms.

Indeed, the epigenome of differentiated cells can be reprogrammed to naïve pluripotency *in vitro*, by fusion with an ESC (Tada et al., 2001) or by forced overexpression of key members of the pluripotency TF network. The latter was initially achieved by retroviral transduction of murine fibroblasts with Oct4, Klf4, Sox2 and cMyc transgenes (Takahashi and Yamanaka, 2006) in FCS+LIF conditions. Induced pluripotent stem cells (iPSCs) have since been generated by a range of methods, from a variety of differentiated tissues and species (reviewed in Ezashi et al., 2016; Raab et al., 2014; Theunissen and Jaenisch, 2014). iPSCs share defining functional characteristics with ESCs (Okita et al., 2007; Wernig et al., 2007), although some concerns have been expressed regarding epigenetic memory of the source tissue in early passages (Ohi et al., 2011; Polo et al., 2010).

Reprogramming of somatic cells to iPSCs is a testament to the power of TFs as specifiers of cellular identity. Perhaps unsurprisingly, many beneficial factors for reprogramming are those also implicated in ESC maintenance (Section 1.2.3). Intriguingly, expression onset of such factors *in vivo* precedes naïve epiblast emergence (Boroviak et al., 2014; Guo et al., 2010), consistent with the notion of naïve pluripotency specification *in vivo* as an active

reprogramming event rather than a passive loss of totipotency. It will be of interest to ascertain to what extent there are parallels between naïve pluripotent identity specification *in vivo* vs during *in vitro* reprogramming.

Oct4 is the only member of the original Yamanaka cocktail which cannot be substituted by other family members, and all reprogramming methods either include exogenous Oct4 or factors/chemicals which activate endogenous Oct4 (Radzisheuskaya and Silva, 2014). The importance of Oct4 in somatic reprogramming was underscored by the recent discovery that CRISPR-targeted epigenetic activation of the endogenous *Oct4* locus is sufficient to trigger fibroblast reprogramming to iPSCs, albeit at only 0.01% efficiency (Liu et al., 2018).

As in ESC maintenance, the signalling environment also plays a key role during the induction of naïve pluripotency. LIF is required in the culture medium for almost all reprogramming methods, while PD and CH enhance reprogramming efficiency (Silva et al., 2008; Sridharan et al., 2009; Tang et al., 2012; Theunissen et al., 2011; Yang et al., 2010). This indicates that 2iLIF components not only maintain the naïve pluripotent identity, but can play an active role in its specification.

1.2.6 Primed pluripotency

In the post-implantation epiblast, the pluripotent compartment has progressed to its primed form (Nichols and Smith, 2009) (Figure 1.1.1A). This is a distinct identity, with markedly different transcriptional, epigenetic and metabolic profiles to naïve pluripotency (Figure 1.2.3). These cells can be captured in culture as epiblast stem cells (EpiSCs) and require stimulation rather than inhibition of Fgf signalling, together with the addition of ActivinA (FA) (Brons et al., 2007; Tesar et al., 2007). Whilst 2iLIF-cultured ESCs are the *in vitro* counterpart of the E4.0–4.5 naïve epiblast, EpiSCs correspond to E7.25–8.0. At this time, gastrulation and germ layer specification are well underway, i.e. the epiblast is already acquiring both positional and identity specification within a rich and polarised signalling environment (Beddington and Robertson, 1999). Transcriptionally, EpiSCs resemble the anterior primitive streak of the late gastrula stage (Kojima et al., 2014; Tsakiridis et al., 2014). Since they express some lineage markers, activate enhancers of developmental regulators and exhibit differentiation biases (Bernemann et al., 2011; Factor et al., 2014; Kojima et al., 2014; Tsakiridis et al., 2014), EpiSCs are considered to be primed for development.

Primed EpiSCs and naïve ESCs share hallmarks of pluripotency: environment-dependent self-renewal or multilineage differentiation *in vitro*, the capacity to form teratomas when injected into adult mice, and expression of core pluripotency TFs Oct4, Sox2 and Nanog. However, EpiSCs do not contribute to adult chimeras following blastocyst-injection (Guo et al., 2009; Tesar et al., 2007). When clumps of EpiSCs are grafted onto post-implantation epiblast stages, integration and initial contribution to all three germ layers is seen after 48 hours of embryo culture (Huang et al., 2012). This indicates functional pluripotency in their corresponding embryonic environment, but it cannot be excluded that EpiSC pluripotency is at the population rather than single-cell level. Furthermore, EpiSCs have passed the window of competence for PGC specification, signifying a significant loss of developmental potential (Hayashi et al., 2011; Ohinata et al., 2009).

Important molecular differences between ESCs and EpiSCs further delineate these as distinct cellular identities (Figure 1.2.3). While there are two active X chromosomes in female naïve cells, one of these has been randomly inactivated in primed epiblast cells and correspondingly in EpiSCs (Guo et al., 2009; Okamoto et al., 2004). Compared to the open ESC epigenetic landscape (Section 1.2.5), EpiSCs exhibit a more closed chromatin conformation (Hassan-Zadeh et al., 2017), many-fold higher levels of DNA methylation (Auclair et al., 2014; Hackett et al., 2013; Veillard et al., 2014; Zylicz et al., 2015), and utilise different enhancer elements even for mutually expressed genes (Buecker et al., 2014; Factor et al., 2014).

In vivo, the epiblast converts to a columnar epithelium upon implantation (Kaufman, 1992). Correspondingly, EpiSCs grow *in vitro* in flat, compact colonies and, unlike clonogenic ESCs, do not survive single-cell dissociation (Brons et al., 2007; Tesar et al., 2007). A metabolic switch also occurs between naïve and primed pluripotent states, with bivalent usage of both glycolysis and mitochondrial respiration in the former while the latter are predominately glycolytic (Zhou et al., 2012).

Whilst ESCs and EpiSCs both express Oct4, Sox2 and Nanog, there are regulatory differences corresponding to those observed *in vivo*. Oct4 expression level is similar, but is driven from its distal vs proximal enhancer in naïve vs primed cells respectively (Buecker et al., 2014; Yeom et al., 1996). The transition from one to the other is gradual and Oct4 expression is maintained throughout implantation (Han et al., 2010). In contrast, Nanog is downregulated during implantation then reexpressed in the prospective primitive streak of gastrulating embryos (Hart et al., 2004). Sox2 expression persists in the epiblast throughout implantation, then becomes anteriorly restricted by mid–late-streak stages (Avilion et al., 2003; Wong, 2015). EpiSCs express both Nanog and Sox2, but at $\leq$20% of ESC-level and there is

some evidence their expression levels are reciprocal within single EpiSCs (Guo et al., 2009; Han et al., 2010; Wong, 2015), recapitulating their opposing gradients in the post-implantation epiblast. The action of the Oct4/Sox2/Nanog network core also differs between naïve and primed pluripotent cells, with distinct genomic targets and binding partners such as Oct6, Otx2 and Sox3 in the primed state (Buecker et al., 2014; Galonska et al., 2015; Tesar et al., 2007).

The transcriptional network governing EpiSC identity is far less well characterised than that of naïve pluripotency. In addition to Oct4/Sox2/Nanog, Utf1, Otx2, Oct6, Sox3 and Zic2/3 (Brons et al., 2007; Corsinotti et al., 2017; Galonska et al., 2015; Iwafuchi-Doi et al., 2012; Tesar et al., 2007), the regulatory network governing primed pluripotency presumably includes additional factors, connected in an unknown network topology. Other markers of EpiSCs include Fgf5, T/Brachyury, Eomes, FoxA2, Lefty, Sox17, Gata4. With the exception of Fgf5, these are considered indicators of differentiation priming/bias rather than intrinsic components of the EpiSC regulatory network, and vary greatly between EpiSC lines (Kojima et al., 2014; Morgani et al., 2017).

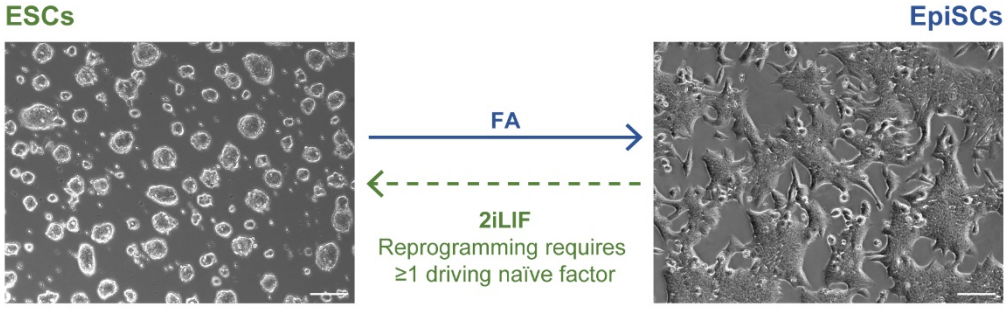
The mechanisms by which FA signals sustain EpiSC identity *in vitro* are also partially understood. Nanog has been identified as a direct but non-essential target of Activin/Smad2/3 signalling in EpiSCs (Greber et al., 2010; Osorno et al., 2012; Vallier et al., 2009). Meanwhile, Fgf/Erk signalling blocks neuroectodermal specification, and supresses naïve factors such as Klf2 (Greber et al., 2010; Yeo et al., 2014). An improved understanding of the EpiSC TF network will be required before we can define how the signalling environment interacts with this to govern the primed pluripotent identity.

Provision of FA causes ESC differentiation into an EpiSC-like state (Guo et al., 2009), recapitulating epiblast progression from Erk-independence to -dependence. Conversely, naïve culture conditions are insufficient to convert EpiSCs back to naïve pluripotency[1] (Figure 1.2.3). This indicates a barrier between the primed and naïve pluripotent identities, in the form of epigenetic restriction, network hysteresis, and/or signal insensitivity. EpiSC reprogramming can be achieved by forced expression of certain TFs, in cooperation with switch to the 2iLIF signalling environment. The first such factor identified was Klf4 (Guo et al., 2009). Since then it has become apparent that almost any single member of the naïve TF network can drive EpiSC reprogramming, including Esrrb, Gbx2, Klf2, Klf4, PStat3, Tbx3 and Tfcp2l1 (Festuccia et al., 2012; Guo et al., 2009; Hall et al., 2009; Martello et al., 2013; Silva et al., 2009; Yang

[1] Except in rare cases at very low efficiency, for a few privileged feeder-derived EpiSC lines or subpopulations thereof (Han et al., 2010). However, such cells likely occupy an earlier identity to start with (Morgani et al., 2017).

et al., 2010; Ye et al., 2013). This hints at a fascinating flexibility in naïve network installation. Whilst single-factor EpiSC reprogramming has so far been inefficient (<1–3%), efficiencies of up to 10% have been reported due to synergy between drivers (Gillich et al., 2012; Yang et al., 2010). It is not yet known whether the TF network interactions are the same during naïve pluripotency induction as opposed to maintenance.

Epigenetic perturbation can also prompt EpiSC reprogramming (Zhang et al., 2016). This suggests that the epigenetic landscape plays an instructive role in reprogramming, albeit via expression modulation of lineage- vs pluripotency-specifying TFs and in the presence of 2iLIF signals. Indeed, 2i has been shown to trigger reconfiguration of Oct4/Sox2/Nanog genomic binding prior to epigenomic remodelling (Galonska et al., 2015), and to upregulate naïve TF network members and known reprogramming drivers such as Klf2 and Tfcp2l1 (Martello et al., 2013; Ye et al., 2013; Yeo et al., 2014). The instructive role of the environment in EpiSC reprogramming is further demonstrated by the limiting nature of Stat3 activation for EpiSC reprogramming (Yang et al., 2010), and the dominance of forced Stat3 activation over antagonistic cues (van Oosten et al., 2012).



Pluripotent state	Naïve	Primed
Culture condition	2iLIF	FA
In vivo counterpart	E4.0-4.5	E7.25-8.0
Functional potential	Pre-implantation chimeras; unbiased; germline transmission	Post-implantation chimeras; biased; very poor PGC formation
Epigenetic profile	XaXa; open chromatin; hypomethylated; use of distal enhancers on average	XaXi; closed chromatin; hypermethylated; use of proximal enhancers on average
TF expression	Oct4, Nanog, Sox2 Esrrb, Fgf4, Gbx2, Klf2/4/5, Rex1, Stat3, Tbx3, Tcl1, Tfcp2l1	Oct4, Nanog (less), Sox2 (less) Fgf5, FoxA2, Gata4, Lefty, Oct6, Otx2, Sox17, T, Utf1, Zic2/3
Metabolism	Bivalent	Glycolytic
Signalling	Jak/Stat3 active; Wnt active; Fgf/Erk inibited; Tgfβ/Smad inactive	Jak/Stat3 inactive; Wnt inhibition beneficial; Fgf/Erk active; Tgfβ/Smad active
Morphology	Compact domed colonies; highly clonogenic	Flat epithelial colonies; low clonogenicity

Figure 1.2.3: Primed and naïve pluripotency

Summary of the differences between naïve and primed pluripotent identities. Representative phase images are shown of ESCs cultured in 2iLIF and EpiSCs in FA. Scale bars: 100 μm.

1.2.7 Human pluripotency

Following the advances in the murine system, human pluripotent cells were first captured *in vitro* 20 years ago, from supernumerary embryos (Thomson et al., 1998). However, despite being derived from pre-implantation blastocysts, conventional human (h)ESCs share defining features with primed murine EpiSCs, such as signalling requisites, flattened morphology, intolerance to culture at clonal density, glycolytic metabolism, proximal Oct4 enhancer usage, epigenetic landscape, and an inactive X chromosome if female (Greber et al., 2010; Nichols and Smith, 2009; Tesar et al., 2007). Heterogeneity and lineage bias are also prevalent in conventional (primed) hESCs (Blauwkamp et al., 2012; Enver et al., 2009; Hough et al., 2014; Osafune et al., 2008; Stewart et al., 2006).

For some time, it remained unclear whether this represented the properties of the human pre-implantation blastocyst, or whether it was a culture-induced state due to suboptimal conditions. Assuming it existed, extensive efforts to capture the putative human naïve pluripotent identity recently bore fruit, in an exciting advancement during the course of my PhD studies. By expression of Klf2+Nanog transgenes and/or modulation of the signalling environment, hESCs were reset to a naïve state that shares molecular attributes and transcriptional governance with murine naïve ESCs (Takashima et al., 2014; Theunissen et al., 2014). Subsequently, naïve ESCs were successfully derived directly from embryos (Guo et al., 2016), with which they share hallmark naïve-specific Mek/Erk independence (Roode et al., 2012) and transcriptional signature (Stirparo et al., 2018). This forms a strong argument for the conservation of naïve–primed pluripotency progression in mammals, although with species-specific features within this overarching framework.

However, we have yet to reach a consensus regarding the optimal culture condition for capture and propagation of human naïve pluripotency (Davidson et al., 2015; Morgani et al., 2017). In part, this is impeded by limited scope for *in vivo* comparison or validation. An improved mechanistic understanding of murine primed-to-naïve identity conversion could instruct ongoing attempts to improve cognate naïve hESCs. Furthermore, functionality of naïve hESCs for *in vitro* multilineage differentiation is currently challenged by protracted timelines and difficulty accessing mature cell types. Since the *in vivo* human naïve epiblast is clearly capable of achieving this in a timely manner, likely our culture methods and differentiation protocols still have much to learn in order to faithfully recapitulate development.

1.3 Experimental study of cell identity transitions

1.3.1 Outstanding questions regarding mechanisms of cell identity transition

Returning to a general consideration of cell identity transitions, execution of a transition necessitates a multitude of molecular changes to occur. The source transcriptional network needs to be dismantled and the destination program installed. Likewise, the epigenome, signal production and sensitivity, metabolism, morphology, substructures and specialised functionalities must be appropriately reconfigured. As outlined in Section 1.1, many of these cellular properties are ultimately dictated by the transcriptional network in charge of a given identity. Therefore, for this thesis I focus on reconfiguration of the transcriptional control network as a defining feature of cell identity transition. Outstanding questions regarding how this is achieved include:

1. How does imposition of one/a few TFs lead to coordination of the destination network?
2. How do external signals interact with the TFs to instruct the destination identity?
3. Are there multiple routes by which a given identity change can occur, or must it always follow the same progression of mechanistic steps?
4. Are there limiting steps for network rewiring?
5. Do cells need a 'licence' to execute a transition?

The specifics of these parameters will undoubtedly differ on a case-by-case basis. Nevertheless, at their core these are fundamental questions of wide interest. In particular, question 3 pertains to key principles for our understanding of multicellular biology as a function of genome-subset usage choreographed over space and time. Returning to the model of cellular identity as an attractor (Sections 1.1.3–4), question 3 can be posed as:

3. Is an identity attractor multidimensional, with multiple ways by which it can be approached, or do transitions follow a set path through an energetic 'valley'?

Empirical evidence supporting theories of cellular identity as a multidimensional attractor was provided in a landmark work by Huang et al., 2005. They showed two transcriptionally distinct routes of promyelocytic HL60 cell differentiation into neutrophils, although they noted some disparity in the resulting neutrophil identities. A limitation for the further understanding of the principles governing cell identity change has been a lack of suitable *in vitro* cell types and of defined, tractable systems to study the transitions occurring between these. Here, I aim to address this by generation and implementation of appropriate *in vitro* models.

1.3.2 Requirements of an appropriate identity transition model

In order to experimentally address the above questions (Section 1.3.1), identity transition models are required with the following attributes:

a. Source and destination identities must be well defined.

b. Source and intermediate cells must be amenable to perturbation.

c. Genetic and signal variables can be controlled and independently manipulated.

d. Responses should be sufficiently rapid to ascribe direct regulatory relationships.

e. For the sake of technical feasibility, an efficient identity change would be preferable.

f. Either we are confident that averages are representative of the intended identity change, or there is a means to isolate productively transitioning cells.

1.3.3 Choice of naïve pluripotency as the destination cell identity

I chose murine naïve pluripotency as the destination identity for the study of cell identity transitions. This provides several advantages: the destination identity is extremely well defined in terms of its molecular signature (Section 1.2), while functional assays such as clonogenic expansion and chimeric contribution leave no doubt as to whether the identity in question has indeed been generated (Figure 1.1.1B).

Its high degree of molecular and functional definition render naïve pluripotency an appropriate destination for consideration of whether identity attractors are multidimensional, conceivably overcoming previous limitations with destination discrepancies (Huang et al., 2005). Furthermore, the defined serum-free culture of naïve pluripotent cells in 2iLIF leaves little room for other identities to propagate, particularly since Erk-signalling is required for most cell types, and this can be further strengthened by implementation of naïve selection reporters (Chambers et al., 2007; Toyooka et al., 2008; Wray et al., 2011). The well characterised homogeneity of 2iLIF cultures lends further confidence to the uniformity of identities propagated therein (Hackett et al., 2017), although of course validation will be essential.

It is intrinsically intriguing to generate the identity that itself has the unbiased potential to form all others of the adult organism. Furthermore, a deeper understanding of murine naïve pluripotent identity specification may improve stem-cell based developmental models, and inform capture and manipulation of cognate pluripotent cells from other species.

1.3.4 Choice of EpiSC reprogramming to naïve pluripotency as the model system

Having set naïve pluripotency as the destination identity, I then had to decide on the source cell type for a tractable model of identity transitions. Reprogramming of somatic cells to iPSCs requires multiple genetic and signal variables to be introduced simultaneously in order to achieve reprogramming, prohibiting causal ascription of changes to individual inputs (Smith, Sindhu & Meissner, 2016). Somatic cell reprogramming is a multi-step process, of which naïve pluripotency establishment is the final transition and occurs at low efficiency.

In stark contrast, reprogramming of EpiSCs requires only one driving naïve factor, combined with defined modulation of the culture signalling environment (Section 1.2.8). Indeed, both EpiSC and naïve iPSC culture regimes are defined, serum- and feeder-free, with a shared N2B27 basal medium and fibronectin coating. The only environmental variables are thus exchange of FA ligands for 2iLIF components, and in certain cases even this is unnecessary (van Oosten et al., 2012). Together with independent manipulation of driving transgene(s), this proffers an exceptional degree of parameter control. Nevertheless, EpiSC reprogramming constitutes a transition between truly distinct identities (Section 1.2.8), and thus provides an appropriate model system in which to study regulatory principles of naïve identity acquisition.

1.3.5 Challenges to overcome

Despite the above advantages, a number of challenges persist in order to generate a tractable system for the mechanistic study of identity change during EpiSC reprogramming to naïve iPSCs. At the start of my doctoral studies, the transcriptional kinetics of naïve network induction during EpiSC reprogramming were unknown. Therefore, I needed to ascertain whether these are/can be sufficiently rapid to support mechanistic study of identity installation (see Chapter 3).

Bulk averages are confounded by the heterogeneous and asynchronous nature of reprogramming. Enrichment or tracing of productive intermediates has been addressed in somatic cell reprogramming (Buganim et al., 2012; O'Malley et al., 2013; Polo et al., 2012). Here, starting instead from EpiSCs, I will need to overcome this limitation by developing methodology to isolate the productive reprogramming population and/or by improving reprogramming efficiencies without compromising the number of variables (see Chapter 4).

Whilst the action of PD and CH small molecule inhibitors is not dependent on intrinsic properties of the cell, meaningful provision of exogenous LIF necessitates competence of the receiving cell to transduce the signal. EpiSCs have an inadequate ability to activate Stat3 in response to LIF, with LIF receptor (LIFR) identified as the limiting component (Yang et al., 2010). However, due to the importance of PStat3 in the naïve pluripotent network (Section 1.2) it should be included as a reprogramming driver. Fortunately, specific and sustained Stat3 activation can be achieved by supplying granulocyte colony stimulating factor (GCSF) to cells expressing GY118F, a chimeric transmembrane receptor (Burdon et al., 1999; Niwa et al., 1998) (Figure 1.2.4). The GY118F extracellular domain is that of the human GCSF receptor, whereas the transmembrane and intracellular domains are based on murine GP130, with a point mutation of tyrosine-118 to phenylalanine. The Y118F mutation prevents Jak-mediated activation of Mek/Erk and PI3K pathways, and blocks Socs3-mediated negative feedback by eliminating the Socs3 binding site on GP130 (Auernhammer et al., 1999; Schmitz et al., 2000). GCSF-stimulation of GY118F is capable of activating Stat3 in EpiSCs and driving their reprogramming to naïve iPSCs (van Oosten et al., 2012; Yang et al., 2010), and thus I will employ this elegant transgene as needed.

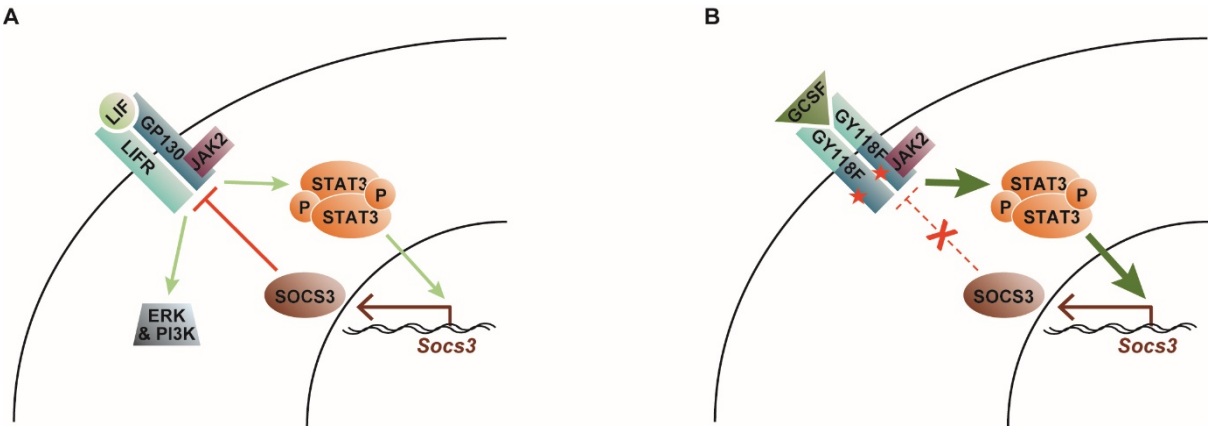


Figure 1.2.4: Stat3 activation by endogenous LIFR vs by transgenic GY118F

Schematic comparing LIF/LIFR and GCSF/GY118F signal transduction pathways.
(A) LIF ligand binds the extracellular domains of LIFR and GP130, triggering a tyrosine kinase signalling cascade that activates Stat3, Mek/Erk and PI3K pathways. Socs3 expression is rapidly induced by PStat3, then binds to GP130-Jak2 complexes. This prevents Stat3 binding to Jak2 and thus inhibits Stat3 phosphorylation and activation, forming a classic negative feedback loop in which PStat3 and Socs3 levels oscillate in antiphase (Kershaw et al., 2013; Schmitz et al., 2000; Yoshiura et al., 2007).
(B) Specific and sustained Stat3 activation can be achieved by supplying GCSF to cells expressing GY118F transgene. The Y118F point mutation prevents stimulation of Mek/Erk and PI3K signalling, and blocks Socs3-mediated negative feedback.

Finally, I considered how best to introduce other single reprogramming factors to EpiSCs. When I started my doctoral studies, reported cases of EpiSC reprogramming had been driven by constitutive overexpression of factors in EpiSCs that would coordinate reprogramming upon switch to 2iLIF. However, given my interest in studying whether there could be multiple routes for transition between two given identities, I was concerned that prolonged expression of naïve factors in EpiSCs prior to the reprogramming assay could compromise the equivalency of the starting identity. Whilst transient transfections overcome this problem, they introduce others such as transfection stress at the critical moment of reprogramming initiation, and challenges in reproducibility. Therefore, I decided to try stably-integrated dox-inducible driver transgenes (see Chapters 3–4).

1.3.6 Summary of aims

I aim to study the principles underpinning cell identity transitions, using reprogramming of EpiSCs to naïve iPSCs as a model. This necessitates development of defined and tractable identity transition models. Using these systems, I seek to address the following biological questions:

How is the naïve pluripotent identity instructed by interplay between transcriptional networks and environmental signals?

Chapter 3, with returns to this theme in Chapters 4–5.

Are there multiple routes by which to transit from EpiSC to naïve iPSC identity?

Chapters 4–7.

CHAPTER 2: MATERIALS AND METHODS

2.1 Tissue culture

2.1.1 Murine tissue culture

ESCs and iPSCs were cultured in N2B27+2i+LIF (2iLIF) unless otherwise indicated. EpiSCs were cultured in N2B27+FGF2+ActivinA (FA). XAV was also added to FA for Chapters 4–7, following evidence that it homogenises primed cultures (Kim et al., 2013; Sumi et al., 2013). Cell lines are of predominately 129 genetic background unless stated otherwise. N2B27 medium comprised 1:1 DMEM/F-12 and Neurobasal (Gibco), 2 mM L-glutamine (Gibco), 1x penicillin-streptomycin (Sigma), 0.1 mM 2-mercaptoethanol (Gibco), 1% B27 (Gibco) and 0.5% N2 (homemade). As required, N2B27 was supplemented with 20 ng/ml murine LIF (homemade), 3 μM CHIR99021 (Chiron; CH) and 1 μM PD0325901 (PD03; PD) (Stewart lab, Dresden), 12.5 ng/ml FGF2 and 20 ng/ml ActivinA (homemade), 6.25 μg/ml XAV 939 (Tocris), 3 μM DMH2 (Tocris), or 0.6 μM LDN193189 (Sigma).

Cells were occasionally cultured in FCS+LIF or KSR+LIF instead, if necessary for the objective of the experiment. FCS+LIF medium contained GMEM (Sigma), 10% fetal calf serum (Gibco), 1x non-essential amino acids (Gibco), 1 mM sodium pyruvate (Sigma), 2 mM L-glutamine (Gibco), 1x penicillin-streptomycin (Sigma), 0.1 mM 2-mercaptoethanol (Gibco), 20 ng/ml murine LIF (homemade), and 10ng/ml BMP4 (Miltenyi Biotec) was supplemented as indicated. KSR medium contained GMEM (Sigma), 10% KnockOut Serum Replacement (Invitrogen), 1% fetal calf serum (Gibco), 1x non-essential amino acids (Gibco), 1 mM sodium pyruvate (Sigma), 2 mM L-glutamine (Gibco), 1x penicillin-streptomycin (Sigma) and 0.1 mM 2-mercaptoethanol (Gibco).

For ESCs and iPSCs, tissue-culture flasks (Falcon) were coated with 0.15% gelatin (Sigma) in PBS (Sigma) and incubated in 7% CO_2. For EpiSC culture and reprogramming experiments, tissue-culture flasks were coated with 10 μg/ml fibronectin (Millipore) in PBS (Sigma) and incubated in 7% CO_2 (Chapter 3) or 7% CO_2 and 5% O_2 (Chapters 4–7). ESCs, iPSCs and EpiSCs were dissociated with accutase (Millipore) during passaging. For optimal performance of EpiSCs, lines were maintained by plating 25000 cells/cm^2 every other day (usually 1:6 split ratio) following gentle accutase treatment for less than 3 minutes at room temperature.

2.1.2 Murine cell transfection

For transgene integration transfections, 1 µg piggyBac (PB) vectors of interest, 0.5 µg PBase expression vector (*CAG.PBase*) and 10 µl Lipofectamine-2000 (Invitrogen) were incubated for 20 min in 500 ml DMEM (Gibco), then applied to 500,000 cells/6well in 3 ml medium for 18 hours. Selection was applied to transfectants for at least 5 passages prior to use: 50–150 µg/ml hygromycin-B (Life Technologies), 0.33–1.00 µg/ml puromycin (ThermoFisher), 20 µg/ml blasticidin (Gibco), 200 µg/ml G418 (Invitrogen) or 100 µg/ml zeocin (Invitrogen) as appropriate. PB-vectors stably integrate into host genome TTAA-sites in the presence of PBase transposase; *CAG.PBase* does not integrate and is lost with passaging (Cambridge Bioscience).

siRNA transfections were performed using RNAiMAX transfection reagent (Invitrogen) and FlexiTube siRNAs against Oct4, Klf2, or AllStars Negative Control (Qiagen) according to the manufacturers' instructions.

2.1.3 EpiSC reprogramming

EpiSCs were plated in FA without selection at a density of 1000/cm^2. For siRNA experiments, 5000/cm^2 was used instead to compensate for transfection toxicity. The following day, reprogramming was induced by medium change to 2iLIF or subset components thereof as indicated. GY118F transgenic receptor was stimulated with 30 ng/ml human GCSF (Peprotech), whereas expression of other transgenes was induced with 1 µg/ml doxycycline (MP Biomedicals).

After 4–7 days, transgene induction was withdrawn and selection applied to select for naïve iPSC colonies. As appropriate: 20 µg/ml blasticidin (Gibco) was used to select for *Rex1::*dGFP.IRES.bsd reporter activity; 1 µg/ml puromycin (ThermoFisher) or 200 µg/ml G418 (Invitrogen) to select for *Oct4::*GFP.IRES.puro or *Oct4::*βgeo; 50–100 µg/ml G418 (Invitrogen) to select for *Nanog::*βgeo.

On day 8–12, 4x images were acquired using CellSens software and an X-51 Olympus microscope system with motorized stage and camera. iPSC colonies with active *Rex1, Nanog* or *Oct4* reporter were counted manually. Unless stated otherwise, reprogramming data presented in this thesis are the mean of 3 biological replicates.

2.1.4 FixedOct4 EpiSCs

FixedOct4 EpiSCs were generated from *Oct4-/βgeo* CAG.Oct4^{wt}.2A.mCherry iPSCs (Radzisheuskaya et al., 2013) by differentiation in N2B27+Fgf2+ActivinA for 10 passages. *Oct4 F/βgeo CAG.EmptyVector* EpiSCs were generated as a control from the same parental line. Reprogramming was conducted in N2B27+2iLIF as above. 200 µg/ml G418 (Invitrogen) was applied to select for endogenous *Oct4* promoter activity after 4 days of reprogramming.

2.1.5 Somatic cell reprogramming from E9.5 chimeras

High contribution E9.5 chimeras were generated following blastocyst injection of *Oct4-/βgeo* CAG.Oct4^{wt}.2A.mCherry iPSCs (Radzisheuskaya et al., 2013). The anterior portion of E9.5 chimeras were manually dissociated then cultured in N2B27- or KSR-based LIFaza or 2iLIF. KSR-medium comprised GMEM (Sigma), 10% KSR (Invitrogen), 1% FCS (Sigma), 1x non-essential amino acids (Gibco), 1 mM sodium pyruvate (Sigma), 2 mM L-glutamine (Gibco), 1x penicillin-streptomycin (Sigma), 0.1 mM 2-mercaptoethanol (Gibco). 5-Azacytidine (aza) was added as indicated (1 µM, Sigma). After 6 days, aza was removed and 2iLIF was applied to all, followed by addition of G418 (400 µg/ml) at day 10. iPSCs could be derived after a period in LIFaza in both N2B27- and KSR-based media.

2.1.6 Human tissue culture

H9 hESCs (WiCell, agreement #17-W0189) and derivate lines thereof were maintained in E8 media (homemade) on tissue-culture plates coated with Geltrex (ThermoFisher) and incubated at 7% CO_2 and 5% O_2. For expansion, hESCs were passaged in small clumps following dissociation with 0.5 mM EDTA (ThermoFisher). For transfection or resetting assays, hESCs were dissociated to single cells using accutase. H9 iK2N.Venus hESC line was a kind gift from Austin Smith (Takashima et al., 2014).

2.1.7 Human cell transfection

For transgene integration transfections, 1.2 µg piggyBac (PB) vectors of interest and 1.2 µg PBase expression vector (*CAG.PBase*) were electroporated into 1 million hESCs using the Neon transfection system (ThermoFisher) according to manufacturer's instructions. Two

pulses of electroporation were provided, each for 2 ms at 1200 V, then hESCs were plated at clonal density with 10μM ROCKi/Y-27632 (Calbiochem). After two days, ROCKi was withdrawn and selection applied until stable cell lines were established: 50 μg/ml hygromycin-B (Life Technologies) for *PB.TetO.GOI.PGK.hph* or *PB.CAG.GY118F.PGK.hph*, and 0.33 μg/ml puromycin (ThermoFisher) for *PB.CAG.rtTA3.PGK.pac*.

2.1.8 Human cell resetting

Resetting of primed H9 iK2N.Venus hESCs towards the naïve state was conducted in N2B27+2iLIF on mitotically inactivated MEFs, as previously described (Takashima et al., 2014). 1 μg/ml doxycycline (MP Biomedicals) or 3 μM DMH2 (Tocris) were applied as indicated. 4x images were acquired using CellSens software and an X-51 Olympus microscope system with motorized stage and camera. Venus+ colony size was analysed in ImageJ using the automated 'Analyse Particles' function.

2.2 Embryo work

2.2.1 Derivation of EpiSCs

To derive Nanog *F/βgeo* EpiSCs, Nanog *F/F* (Chambers et al., 2003) and Nanog *+/βgeo* (Mitsui et al., 2003) mice were crossed and EpiSCs were derived from resultant E6.5 embryos. Epiblasts were manually dissected from extra-embryonic tissues and plated on fibronectin-coated plates in FGF2+ActivinA medium. Nanog genotyping was performed on extra-embryonic tissues and confirmed on downstream EpiSC cultures.

To derive *Rex1*-reporter EpiSCs, *Rex1*::dGFP.IRES.bsd homozygous 129 studs (Kalkan et al., 2017) were crossed with wild-type 129 females and EpiSCs were derived from resultant E6.5 embryos. Epiblasts were manually dissected from extra-embryonic tissues and plated on fibronectin-coated plates in FGF2+ActivinA+XAV medium.

After 5–7 days of culture, regions of the explant exhibiting EpiSC morphology were manually passaged to a fresh plate. Subsequent passages were performed using accutase (Millipore).

2.2.2 Embryo culture and analysis

Wild-type 129 mice were crossed and embryos flushed from oviducts at 2.75 dpc using M2 medium (Millipore). Embryos were subsequently incubated in Blast medium (Origio) and periodically inspected. At cavitation onset, embryos were randomly divided into Blast medium supplemented either with 3 μM DMH2 (Tocris) or 1:1000 DMSO, cultured for a further 36 h then fixed in 4% paraformaldehyde (Sigma) in PBS for 15 min at room temperature.

Embryos were permeabilised in 0.25% Triton X-100 (Sigma) in PBS for 30 min, then blocked in 3% donkey serum (Sigma), 0.1% BSA (Sigma) and 0.01% Tween-20 (Sigma) for 30 min at room temperature. Embryos were incubated overnight at 4°C in blocking buffer with the following primary antibodies: Klf4 (1:300, goat pAb, R&D); Nanog (1:300, rat mAb, eBioscience); Oct4 (1:300, rabbit mAb, Cell Signaling). The following day, washes were performed in blocking buffer. AlexaFluor secondary antibodies (Life Technologies) were used against the appropriate species at 1:1000 in blocking buffer. Embryos were gradually acclimatised then mounted in Vectashield with DAPI (Vector Laboratories).

Images were taken with an Eclipse Ti Spinning Disk confocal microscope (Nikon) equipped with an Andor Revolution XD System at 40x magnification. Presented images are maximum intensity projections of Z-stack slices processed with ImageJ.

Staining quantification was carried out with Imaris: nuclei were identified in the DAPI channel and the fluorescence of each other channel recorded; cells were assigned to the trophectoderm lineage based on position (outer cells), while remaining cells were classified as inner cell mass. There was no significant loss of fluorescence in the Z-dimension and no batch effect was detected in the stainings, so no further normalisation was deemed necessary.

2.2.3 Blastocyst injection

Chimeras were generated by William Mansfield, using standard microinjection methodology and host blastocysts of strain C57BL/6 (black). All animal work was performed in accordance with Home Office guidelines and regulations at the University of Cambridge, UK.

2.3 Flow cytometry

Cell sorting was performed using a MoFlo Legacy Cell Sorter (Beckman) or an S3 Cell Sorter (BioRad). dGFP was excited using a 488 nm laser and detected using a 530/30 filter. *Rex1+/dGFP* EpiSCs and ESCs were used to determine negative and positive dGFP gates respectively.

After sorting of reprogramming intermediates, number of iPSC colonies are quantified relative to the number of ESC colonies, because replating of sorted ESCs provides a control for cell death due to the stress of sorting. ESCs already stably occupy the destination naïve pluripotent identity, and are thus the appropriate functional control.

2.4 Immunohistochemistry

Cells were fixed for 10 min with 4% paraformaldehyde (Sigma) in PBS, permeabilised in 0.4% Triton X-100 (Sigma) in TBS, then blocked in 5% donkey serum (Sigma) and 0.1% Triton X-100 in TBS. Samples were incubated overnight at 4°C in blocking buffer with the following primary antibodies: Klf4 (1:300, goat pAb, R&D); Nanog (1:300, rat mAb, eBioscience); Oct4 (1:300, rabbit mAb, Cell Signaling); Oct4 (1:100, mouse mAb, Santa Cruz); PSmad1/5 (1:100, rabbit mAb, Cell Signaling). The following day, washes were performed with 0.1% Triton X-100 in TBS, and samples incubated with Dapi and AlexaFluor secondary antibodies against the appropriate species at 1:1000 (Life Technologies).

Samples were imaged using an Eclipse Ti Spinning Disk confocal microscope (Nikon) equipped with an Andor Revolution XD System at 40x or 60x magnification, or using a Leica DMI6000 microscope at 20x. Images were processed with ImageJ. Presented images are maximum intensity projections of Z-stack slices.

2.5 Histology

E9.5 embryos were fixed for 4 h at 4°C with 4% paraformaldehyde (Sigma) in PBS, gradually adjusted to 20% sucrose over 2 days, then mounted in OCT and snap frozen on liquid nitrogen. 8 µm cryosections were taken then stored at -80°C. Following rehydration in PBS, sections were permeabilised in 0.25% Triton X-100 (Sigma) in PBS, then blocked in 5% donkey serum (Sigma) and 0.1% Triton X-100 in PBS. Sections were incubated overnight at 4°C in blocking

buffer with the following primary antibodies: Nanog (1:300, rat mAb, eBioscience); Oct4 (1:300, rabbit mAb, Cell Signaling); Oct4 (1:300, goat pAb, Santa Cruz); Sox1 (1:300, rabbit pAb, Cell Signaling); Sox2 (1:300, rat mAb, eBioscience). The following day, washes were performed with 0.1% Triton X-100 in TBS, and samples incubated with Dapi and AlexaFluor secondary antibodies against the appropriate species at 1:1000 (Life Technologies).

Sections were imaged using a Zeiss ApoTome microscope at 20x then tiled. After imaging, H&E histological staining was performed on the same sections according to standard methodologies. These sections were then re-imaged in the same pipeline.

2.6 Western blotting

Cells were lysed in RIPA buffer (Sigma) containing Complete-ULTRA protease-inhibitor and PhoStop phosphatase-inhibitor cocktails (Roche), and sonicated with Bioruptor200 (Diagenode) at high frequency, alternating 30 s on/off for 3 min. SDS-PAGE electrophoresis was performed using Bolt 10% Bis-Tris Plus gels (ThermoFisher) in a Novex MiniCell (ThermoFisher). Protein transfer was performed using the semi-dry iBlot2 system (ThermoFisher) and iBlot Transfer Stacks (ThermoFisher).

The following primary antibodies were used: Esrrb (1:1000, mouse mAb, R&D); Klf4 (1:1000, goat pAb, R&D Systems); Nanog (1:5000, rabbit pAb, Bethyl Laboratories); Oct4 (1:1000, rabbit mAb, Cell Signaling); P-Y705-Stat3 (1:1000, rabbit mAb, Cell Signaling #9145); total Stat3 (1:1000, rabbit pAb, Cell Signaling); αTubulin (1:10000, mouse mAb, Abcam).

Detection was achieved using HRP-linked secondary antibodies at 1:10000 against the appropriate species (GE Healthcare) and ECL Plus Western Blotting Detection System (GE Healthcare).

Membranes were stripped between PStat3 and total Stat3 blots (stripping buffer and protocol as described by Abcam).

2.7 Genotyping

Genomic DNA was extracted by incubation at 95°C for 20 min in 25 mM NaOH + 0.2 mM EDTA, followed by addition of an equal volume of 40 mM Tris HCl and vortexing.

2.7.1 Nanog genotyping

Genotyping to distinguish between Nanog *F/+,* Nanog *F/βgeo* and derivative Nanog *GFP/βgeo* EpiSCs was conducted using Taq DNA polymerase (Qiagen) and the following thermocycler program:
94°C for 5 min; 35 cycles of 94°C for 10 s, 60°C for 20 s, 72°C for 60 s; then 72°C for 3 min.

An equal mix of 3 primers was used:
βgeo AATGGGCTGACCGCTTCCTC
S5 ACCTCAGCCTCCAGCAGATG
A53 CAGAATGCAGACAGGTCTACAGCCCG.

An 800 bp product is amplified from *Flox* (*F*) and wild-type Nanog alleles, whereas the *βgeo* allele yields a 600 bp product.

2.7.2 Oct4 genotyping

Genotyping to distinguish between *Oct4-/βgeo*, *Oct4 F/βgeo* and *Oct4+/+* cells was conducted using Taq DNA Polymerase (Qiagen) and the following thermocycler program:
95°C for 3 min; 30 cycles of 94°C for 15 s, 60°C for 30 s, 72°C for 60 s; then 72°C for 10 min.

Primer GAGCTTATGATCTGATGTCCATCTCTGTGC binds in the *Oct4* final intron, which is present in both wild-type and *Flox* (*F*) alleles.
Primer GGGCTGACCGCTTCCTCGTGCTTTACG binds in the *βgeo* allele.
Primer GCCTTCCTCTATAGGTTGGGCTCCAACC binds 3' downstream of Oct4 and is common to all alleles.

2.8 Plasmids and cloning

Existing piggyBac (PB) vectors were used to drive constitutive expression of: Nanog (*PB.CAG.Nanog.PGK.hph*); GY118F (*PB.CAG.dsRed.IRES.hph.CAG.GY118F*), coding sequence as described in Niwa et al., 1998; Klf4 (*PB.CAG.Klf4.IRES.zeo*); Esrrb (*CAG.Esrrb.IRES.zeo);* Control (*CAG.empty.PGK.hph*). Gateway cloning (Invitrogen) was used to re-insert GY118F to *PB.CAG.GY118F.PGK.hph* when dsRed would confound multichannel IF.

TetO.FLBioNanog.UbiP.rtTA3 was amplified from a previously described dox-inducible Nanog construct [S26] using Phusion Taq polymerase (New England Biolabs). *TetO* denotes the operator/promoter and *rtTA3* encodes the dox-controlled transactivator of the 'Tet-On' system [S27], *FLBio*Nanog codes for FLAG- and biotin-tagged murine Nanog, and the ubiquitous promoter *UbiP* drives constitutive expression of *rtTA3*. *TetO.FLBioNanog.UbiP.rtTA3* was then inserted into PB *bsd* destination vector by Gateway cloning (Invitrogen). To ensure that *FLBio*Nanog was expressed only in the presence of dox, the constitutive *CAG* promoter was removed from the destination vector by restriction digestion and re-ligation. The resulting iFLBioNanog construct confers constitutive blasticidin resistance, allowing selection of successfully transfected cells irrespective of dox presence.

For other dox-inducible transgenes, PB *hph* destination vector was modified to contain TetO rather than CAG promoter (kind gift from Aliaksandra Radzisheuskaya). Genes of interest (Esrrb, Klf2, Klf4, Klf5, Nanog, Tfcp2l1) were cloned into the resultant vector using Gateway technology (Invitrogen) according to manufacturer's instructions, producing inducible expression vectors in the form *PB.TetO.GOI.PGK.hph*. Constitutive rtTA3 expression vector was generated in the same manner, but retaining CAG instead of TetO promoter and thus yielding *PB.CAG.rtTA3.PGK.pac* for co-transfection.

2.9 RT-qPCR

Total RNA was extracted using RNeasy kits, according to manufacturer's spin protocol, including on-column DNaseI digest (Qiagen). cDNA was produced using the SuperscriptIII VILO cDNA synthesis kit, following the recommended protocol. From bulk cultures, 0.2–1.0 µg of input RNA was used. Less RNA was used from sorted samples, corresponding to that extracted from 150–10000 cells. cDNA dilution ratios were adjusted accordingly. RT-qPCR reactions were performed using StepOnePlus Real Time PCR System with recommended

thermocycler settings (Applied Biosystems) and TaqMan Fast Universal PCR Master Mix (Applied Biosystems). Gene expression relative to Gapdh in each well was determined using FAM-labelled TaqMan assay probe together with VIC-labelled Gapdh probe (Applied Biosystems). Unless stated otherwise, RT-qPCR data presented in this thesis are the mean of 3 technical replicates.

Table 1: Applied Biosystems TaqMan RT-qPCR assays

Gene	Probe ID
Gapdh	4352339E
Esrrb	Mm00442411_m1
Fgf5	Mm00438918_m1
Klf2	Mm01244979_g1
Klf4	Mm00516104_m1
Klf5	Mm00456521_m1
Nanog	Mm02384862_g1
Nr0b1	Mm00431729_m1
Nr5a2	Mm00446088_m1
Oct4	Mm00658129_gH
Socs3	Mm01249143_g1
Sox2	Mm03053810_s1
Tfcp2l1	Mm00470119_m1
Rex1	Mm03053975_g1
hGapdh	4326317E
hPou5f1	Hs00232708_m1
hSox2	Hs01053049_s1
hStella	Hs01931905_g1
hTfcp2l1	Hs00232708_m1

Table 2: Applied Biosystems TaqMan RT-qPCR custom assays

Gene	Forward primer (5'–3')	Reverse primer (5'–3')	Probe (FAM 5'–3' MGB)
K4eRNA	AGGCTTTGGCTGGCTGATAA	CTGTCTCCATAGGTACTGACTTCCT	CCCAGCTCAGTAATTG
Retroviral Klf4	TGGTACGGGAAATCACAAGTTTGTA	GAGCAGAGCGTCGCTGA	CCCCTTCACCATGGCTG
Retroviral cMyc	TGGTACGGGAAATCACAAGTTTGTA	GGTCATAGTTCCTGTTGGTGAAGTT	CCCTTCACCATGCCCC

2.10 ChIP

2.10.1 ChIP procedure

Nanog ChIP was performed as follows: cells (10×10^6 for each sample) were fixed for 10 min in 1% formaldehyde, washed with ice-cold PBS and incubated for 10 min in lysis buffer 1 (50 mM HEPES at pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP40 and 0.25% Triton X-100) and then for 10 min in lysis buffer 2 (10 mM Tris at pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA). Nuclei were pelleted, resuspended in shearing buffer (1% SDS, 10 mM EDTA and 50 mM Tris at pH 8.0) and sonicated to obtain an average DNA fragment size of 500 base pairs. Lysates were diluted 1:10 in dilution buffer (50 mM Tris-HCl at pH 8.0, 167 mM NaCl, 1.1% Triton X-100 and 0.11% Na deoxycholate) and pre-cleared for 2 h at 4°C with Dynabeads Protein G magnetic beads (Life Technologies) that were pre-incubated with isotype IgG antibody. The chromatin was then incubated overnight at 4°C with 2 µg of rabbit polyclonal antibody against Nanog (Bethyl Laboratories) or an isotype IgG control (Santa Cruz Biotechnology). Lysates were then incubated for 1 h at 4°C with blocked Dynabeads magnetic beads, and the beads were washed twice in wash buffer 1 (50 mM Tris-HCl at pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na deoxycholate and 0.5 mM EGTA), once in wash buffer 2 (50 mM Tris-HCl at pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na deoxycholate and 0.5 mM EGTA), once in wash buffer 3 (50 mM Tris at pH 8.0, 250 mM LiCl, 0.5% Na deoxycholate, 0.5% NP40, 1 mM EDTA and 0.5 mM EGTA) and twice in wash buffer 4 (50 mM Tris at pH 8.0, 10 mM EDTA and 5 mM EGTA). Chromatin was eluted for 30 min at 65°C in elution buffer (1% SDS and 0.1 M $NaHCO_3$). Samples were incubated overnight at 65°C to reverse the crosslinking and purified using the QIAquick PCR Purification kit (Qiagen). Chromatin was analysed by Fast SYBR Green RT-qPCR (Applied Biosystems). Enrichment was calculated relative to the Input sample.

Table 3: ChIP primers

Region	Forward primer (5'–3')	Reverse primer (5'–3')
Socs3 proximal	GAAAAGGCTTGAGGGTCGGA	CGGGCCTGGAATGTCAAACT
Socs3 distal	TCAGGAGTCCCTGTGCTCTAA	GGCAGACGGGTCTACTTTGAA
Negative control	CTGGGCTTGCAGCTTAGG	AGAGACCTGGCTGAGGATGAC

2.10.2 Analysis of published ChIP-seq datasets

The following published ChIP-seq datasets were used to examine TF binding and histone modifications at genomic regions of interest in ESCs: Nanog (accession number GSE11724); PStat3 (GSE11431); Med12 (GSE22562); RNA PolII (GSE23943); H3K27ac (GSE24164); H3K36me3 (GSE23943); H3K4me3 (GSE23943) and H3K27me3 (GSE23943). Raw ChIP-seq datasets were downloaded from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/), realigned to the mouse mm9 genome with Bowtie software (http://bowtie-bio.sourceforge.net), and peaks were called with MACS software (http://liulab.dfci.harvard.edu/MACS) using default parameters. 'K4-K36' domains were predicted from H3K4me3 and H3K36me3 patterns by Guttman et al., 2009.

2.11 Microarray

Amplification and labelling of RNA were performed using the TotalPrep-96 RNA Amplification Kit for the Illumina platform (Ambion). Subsequent hybridization, staining and scanning were performed according to the Whole Genome Gene Expression Direct Hybridization Guide on the MouseWG-6 v2.0 Expression BeadChip (Illumina). Data were loaded into the R package lumi (Du et al., 2008) and then divided into subsets to be analysed. The data were transformed using Variance Stabilization (Lin et al., 2008) and normalised using quantile normalization. Comparisons were performed in the R package limma (Smyth, 2004) and the results were corrected using the False Discovery Rate. The analysis employed a 5% confidence interval.

2.12 RNAseq

2.12.1 Bulk RNAseq library preparation

Sequencing libraries were prepared according to the SmartSeq2 protocol (Picelli et al., 2014) with the following amendments: purified RNA was used diluted to 5 ng/µl; ERCC spike-ins (Invitrogen) were added at 1 µl of 1:10000 dilution per 5 ng; 13 cycles of amplification were used to obtain cDNA (rather than 21 cycles used for single cells). Nextera XT reactions were scaled-down by half, using 0.4 ng cDNA input per reaction. Pooled libraries were sequenced on the Illumina HiSeq 4000 (paired-end 150 bp reads). Each datapoint in a bulk RNAseq timecourse is a single replicate, to allow all samples to be sequenced together in the same lane to avoid technical biases. Trends were independently confirmed by RT-qPCR.

2.12.2 scRNAseq library preparation

Single cells were index-sorted individually by FACS (BD Influx 5) into wells of a 96-well PCR plate containing lysis buffer on days 2, 3 and 4 following reprogramming induction. dGFP+ DAPI- cells were sorted for all reprogramming, ESC and iPSC samples. dGFP- DAPI- cells were sorted for EpiSCs. scRNAseq was performed as previously described (Nestorowa et al., 2016; Picelli et al., 2014; Wilson et al., 2015). The Illumina Nextera XT DNA kit was used to prepare libraries. Pooled libraries were sequenced on the Illumina HiSeq 4000 (single-end 125 bp reads). Samples from all cell lines were included in each sequencing lane, to control for technical lane effects (Figure 5.2.3A).

2.12.3 RNAseq alignment

GENCODE M12 mouse gene annotation from Ensembl release 87 was used for read alignment (Yates et al., 2016) and splice-junction donor/acceptor overlap settings were tailored to the read length of each dataset. Alignments to gene loci were quantified with HTSeq-count (Anders et al., 2015) based on annotation from Ensembl release 87. Sequencing libraries with fewer than 500,000 mapped reads were excluded from subsequent analyses. Read distribution bias across gene bodies was computed as the ratio between the total reads spanning the 50^{th} to the 100^{th} percentile of gene length, and those between the first and 49^{th}. Samples with ratio >2 were not considered further.

2.12.4 Published embryo scRNAseq datasets

Sequencing data of single-cell mouse embryo profiling studies SRP110669 (Mohammed et al., 2017; E3.5, E4.5, E6.5) and SRP020490 (Deng et al., 2014; compacted morula) were downloaded from the European Nucleotide Archive (Toribio et al., 2017) and aligned as above.

2.12.5 Transcriptome analysis

Principal component and cluster analyses were performed based on $\log_2$ FPKM values and were computed with the Bioconductor packages *DESeq2* (Love et al., 2014), *Sincell* (Juliá et al., 2015) or *FactoMineR* (Lê et al., 2008) in addition to custom scripts. Default parameters were used unless otherwise indicated. For global analyses, genes that registered zero counts

in all single-cell samples were omitted. Euclidean distance and average agglomeration methods were used for cluster analyses unless otherwise indicated. t-SNE analysis was computed using Rtsne R package (Krijthe, 2015) with default parameters. *k*-means hard clustering was performed using the Mfuzz R package (Futschik and Carlisle, 2005) and the optimal number of *k* clusters were selected using the elbow method.

2.12.6 Selection of high-variability genes

Genes exhibiting the greatest expression variability (and thus contributing substantial discriminatory power) were identified by fitting a non-linear regression curve between average $\log_2$ FPKM and the square of coefficient of variation. Indicated specific thresholds were applied along the *x*-axis (average $\log_2$ FPKM) and *y*-axis (CV^2) to identify the most variable genes.

2.12.7 Quadratic programming

Fractional identity of single cells was computed using the R package DeconRNASeq (Gong and Szustakowski, 2013). This package utilises quadratic programming to estimate the proportion of distinctive cell types. The average expression of bulk RNAseq samples (EpiSCs or iPSCs, Fig. S3C) or the average expression of scRNAseq samples (EpiSCs, iPSCs or embryo stages, Fig. 3D) were used as signature datasets. The fraction of identity between single cells and the signature datasets was computed using the whole transcriptome. Reprogramming pseudotimes for single cells were assigned by ordering cells based on their fraction of similarity to EpiSCs (origin) and iPSCs (destination) (Fig. S3C), as previously described (Treutlein et al., 2016).

2.12.8 LOESS regression

Smooth curve local regression (LOESS) lines were fitted to scatter plots of $\log_2$ FPKM vs pseudotime using the R stats package (R Core Team, 2016). Smoothness parameter of 1/3 and 2 degrees of local polynomial were used for curve fitting.

CHAPTER 3: NANOG AMPLIFIES STAT3 ACTIVATION AND THEY SYNERGISTICALLY INDUCE THE NAÏVE PLURIPOTENT PROGRAM

3.1 Introduction

As introduced in Section 1.2.3, the core pluripotency network comprises the TFs Nanog, Oct4 and Sox2. In the naïve pluripotent state, active PStat3 integrates with this network downstream of a LIF-stimulated tyrosine kinase signalling cascade (Chen et al., 2008). LIF is included in all optimal culture conditions for induction or maintenance of naïve cells, including during somatic cell reprogramming. For EpiSC reprogramming, Stat3 activation is limiting (Yang et al., 2010) and, notably, Stat3 signalling alone is sufficient to drive reprogramming in the complete absence of additional culture requisites (van Oosten et al., 2012).

In the absence of maternal Stat3, *Stat3-/-* embryos fail to form a naïve pluripotent epiblast (Do et al., 2013). This classifies Stat3 as a member of the small group of genes essential for the establishment of embryonic naïve pluripotency, along with Nanog, Oct4 and Sox2 (Avilion et al., 2003; Nichols et al., 1998; Silva et al., 2009). Oct4 and Sox2 have distinct biological properties from Nanog and Stat3: despite being essential for the maintenance of pluripotent cells, their overexpression does not enhance ESC self-renewal and instead causes loss of ESC identity (Figure 3.1.1) (Kopp et al., 2008; Niwa et al., 2000; Radzisheuskaya et al., 2013). In contrast, forced expression of Nanog and increased Stat3 activation are each reported to enhance pluripotent cell self-renewal (Chambers et al., 2003; Matsuda et al., 1999). There are several other pluripotency-associated genes that promote ESC maintenance (Section 1.2.3). However, unlike Nanog and Stat3, these genes are not required *in vivo* for the establishment of the naïve pluripotent epiblast (Figure 3.1.1).

Nanog and PStat3 thus hold a special position in the naïve TF network, as rare factors that are both required for naïve pluripotency establishment and beneficial for its maintenance if elevated (Figure 3.1.1). Nanog is a nuclear TF by default, whereas Stat3 activation is dependent on the presence and transduction of exogenous LIF signal. Therefore, the relationship between Nanog and LIF/Stat3 signalling is an important and appropriate model in which to investigate the interplay between TFs and signals for naïve identity regulation.

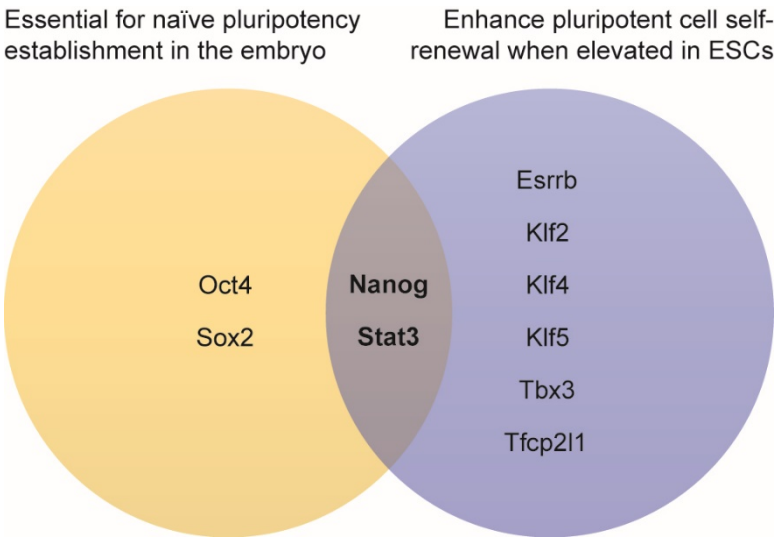


Figure 3.1.1: Nanog and PStat3 hold a special status amongst naïve pluripotency TFs

Venn diagram depicting which naïve TFs are essential for naïve pluripotency establishment in the embryo, and/or which enhance pluripotency maintenance when elevated in ESCs.

It is not known whether Nanog and LIF/Stat3 input to the naïve TF network in an independent or concerted manner. However, in the literature there are several hints at a functional relationship between Nanog and PStat3. Nanog was initially identified due to its ability to support ESC self-renewal in the absence of LIF (Chambers et al., 2003; Mitsui et al., 2003). Furthermore, Nanog overexpression can induce naïve pluripotency in minimal culture conditions if exogenous LIF is supplied (Theunissen et al., 2011). Nanog overexpression and forced Stat3 activation demonstrate hitherto unexplained functional synergy during EpiSC reprogramming (Yang et al., 2010).

Therefore, in Chapter 3 I investigate whether Nanog and LIF/Stat3 signalling are mechanistically linked by assessing their relationship in pluripotent cells and during induction of a naïve pluripotent program. Although Nanog and Stat3 are essential for embryonic pluripotency establishment and ESC derivation, they promote but are not required for *in vitro* ESC maintenance (Chambers et al., 2007; Ying et al., 2008). This permits the study of *Nanog-/-* and *Stat3-/-* ESCs and suggests differences between network requirements for naïve pluripotency establishment and maintenance.

3.2 Results

Please note: certain datasets are not included in this thesis, due to their prior usage in my Bachelors dissertation (University of Cambridge, Biochemistry Part II, 2012). Where conceptually necessary, references are instead provided to the corresponding figures in the resultant publication (Stuart *et al* 2014, Current Biology), in the format 'CB Figure X'. This paper can be found in the Appendix after page 215, and its text is my own original writing.

3.2.1 Nanog amplifies Stat3 phosphorylation and suppresses Socs3 transcription

I first investigated the effect of Nanog on Stat3 activation in ESCs, since the naïve pluripotency network is established and functional in this cellular context. Wild-type (WT) and constitutively Nanog-overexpressing (NanogOE) ESCs were harvested following LIF-induction timecourses. Western blotting revealed that higher Nanog increased PStat3 level and enhanced ESC sensitivity to LIF stimulation (Figure 3.2.1A, see also CB Figure 1A–B).

To explain how Nanog drives PStat3 elevation without directly binding the Stat3 protein or gene (Marson et al., 2008; Torres and Watt, 2008), I examined the effect of Nanog on components of the LIF/Stat3 signalling pathway in ESCs. By RT-qPCR, no correlation was reliably found between Nanog and transcript levels of positive signal transducers Lif, Lifr, Gp130, Jak2 and Stat3 (my Bachelors dissertation), despite Nanog binding to Lif, Lifr and Gp130 gene regulatory sequences in ESCs according to published ChIP-seq data (Marson et al., 2008). However, Nanog also binds the Socs3 gene, which is a negative regulator of Stat3 activation (Kershaw et al., 2013; Schmitz et al., 2000). This prompted my hypothesis that Nanog represses Socs3 transcription.

Since Socs3 transcription is upregulated by PStat3 to form a classic negative feedback loop (Schmitz et al., 2000; Yoshiura et al., 2007), determining whether Nanog overexpression causes Socs3 repression is obfuscated by the effect of PStat3 on Socs3. To disentangle the opposing yet interconnected influences of Nanog and PStat3 on Socs3 expression, I designed experiments in which the PStat3 level is not significantly influenced by Nanog. ESCs can be maintained without exogenous LIF by 2i medium as introduced in Section 1.2.2 (Ying et al., 2008). In the absence of LIF-stimulated Socs3 activation, the effect of Nanog on Socs3 transcription can be assessed. From RT-qPCR analysis of wild-type, *Nanog-/-* and NanogOE ESCs cultured in 2i, strong negative correlation was evident between Nanog and Socs3

expression levels (Figure 3.2.1B). Furthermore, in the absolute absence of PStat3 in *Stat3-/-* ESCs, those constitutively overexpressing Nanog exhibited a lower Socs3 level (Figure 3.2.1B).

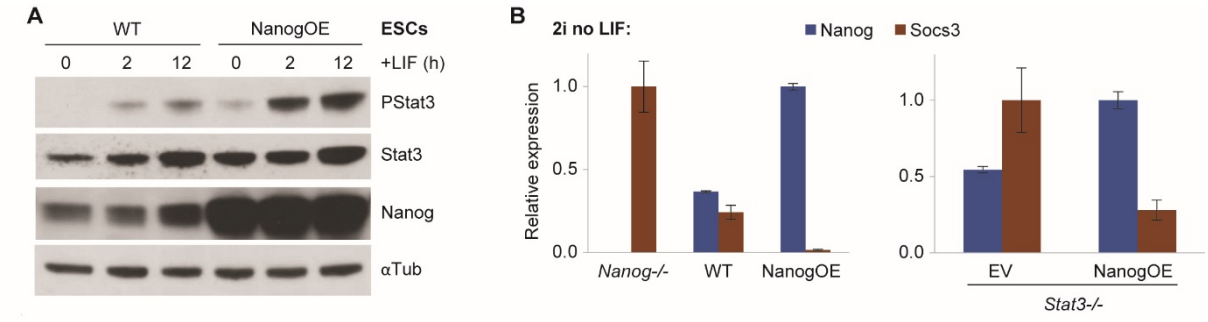


Figure 3.2.1: Nanog amplifies Stat3 phosphorylation and negatively correlates with Socs3

(A) Western blot analysis of PStat3 and total Stat3 protein expression in wild-type (WT) and constitutively Nanog-overexpressing (NanogOE) ESCs in response to LIF stimulation for 0, 2 and 12 hours. LIF was withdrawn from FCS culture for 36 hours prior to re-addition, and selection for pluripotent cells was maintained throughout. Positive feedback of PStat3 induction on Stat3 and Nanog expression was observed, in accordance with previous work (Bourillot et al., 2009; Martello et al., 2013; van Oosten et al., 2012).
(B) RT-qPCR analysis of Nanog and Socs3 expression in ESCs cultured in 2i without LIF for at least 7 days, relative to Gapdh and normalised to the highest values. Mean ± SD (n=3) is shown. NanogOE: constitutive Nanog-overexpression. EV: empty vector.

To permit Nanog perturbation within a single cell line, I generated a doxycycline (dox)-inducible Nanog transgene (iNanog). I amplified *TetO.FLBioNanog.UbiP.rtTA3* from a previously described dox-inducible Nanog construct (Fidalgo et al., 2012). *FLBioNanog* codes for FLAG- and biotin-tagged murine Nanog. *TetO* denotes the operator/promoter and *rtTA3* encodes the dox-controlled transactivator of the 'Tet-On' system (Gossen et al., 1995). The ubiquitous promoter *UbiP* drives constitutive expression of *rtTA3*. *TetO.FLBioNanog.UbiP.rtTA3* was then inserted into piggyBac (PB) blasticidin-resistant destination vector by Gateway cloning. To ensure that *FLBioNanog* was expressed only in the presence of dox, I removed the constitutive *CAG* promoter from the destination vector by restriction digestion and re-ligation. The resulting iFLBioNanog (iNanog) construct confers constitutive blasticidin resistance, allowing selection of successfully transfected cells irrespective of dox presence. Its efficacy and lack of leakage are confirmed in Figure 3.2.2.

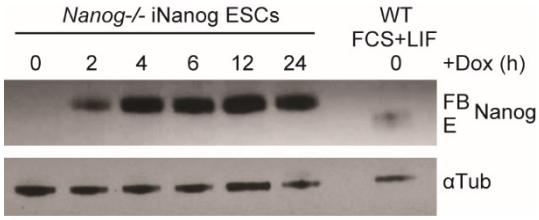


Figure 3.2.2: Validation of dox-inducible Nanog transgene

Western blot of Nanog protein expression in *Nanog-/-* iNanog ESCs following dox induction in FCS+LIF, compared to wild-type (WT) ESCs (N.B. endogenous Nanog expression level is much lower in FCS+LIF than in 2iLIF). FLAG- and biotin-tagged (FB) Nanog expressed from iNanog was detected at a higher molecular weight than endogenous (E) Nanog, as expected.

To independently manipulate Nanog and PStat3 levels within a single cell line, iNanog can be used to induce Nanog on-demand. However, simple LIF-induction of PStat3 could be influenced by whether or not Nanog was co-induced, due to the induction of Socs3 by LIF/Stat3 vs the negative correlation between Socs3 and Nanog (Figure 3.2.1). To solve this problem, I engineered *Nanog-/-* ESCs containing both iNanog and GY118F (iPStat3) transgenes. As introduced in Section 1.3.5, Stat3 can be specifically activated by GCSF-stimulation of the GY118F receptor transgene, which, crucially, is insensitive to repression by Socs3 (see Figure 1.2.4 for schematic). Thus, when GCSF/GY118F are used in the absence of LIF, Socs3 downregulation by Nanog should have little or no effect on PStat3 levels: nearly all Stat3 activation will be attributable to Socs3-insensitivie GY118F rather than Socs3-sensitive LIFR-GP130 (Burdon et al., 1999; Niwa et al., 1998). Therefore, in *Nanog-/-* iNanog+iPStat3 cells in 2i, Nanog and/or PStat3 induction are independent variables.

In 2i without LIF, I induced Nanog expression and Stat3 activation separately and in combination by dox/GCSF respectively. Induction of Nanog alone caused Socs3 repression, while Socs3 induction in response to Stat3 activation was reduced by 60% when Nanog was also induced (Figure 3.2.3A). This suggests that Socs3 repression is a mechanism by which Nanog augments LIF signal transduction in ESCs, resulting in higher levels of active PStat3. The same trends were observed in *Nanog-/-* iNanog+iPStat3 EpiSCs in standard FA conditions (Figure 3.2.3A), demonstrating that Nanog-mediated Socs3 repression is not restricted to ESCs, and may be of functional relevance during Nanog-driven reprogramming.

In order to test whether Nanog could directly repress Socs3 transcription, Nanog-ChIP was performed 3 hours after dox-induction of Nanog in a null ESC background. Induction of Nanog expression rapidly led to enriched Nanog binding at the *Socs3* gene in ESCs, consistent with direct transcriptional regulation (Figure 3.2.3B–C).

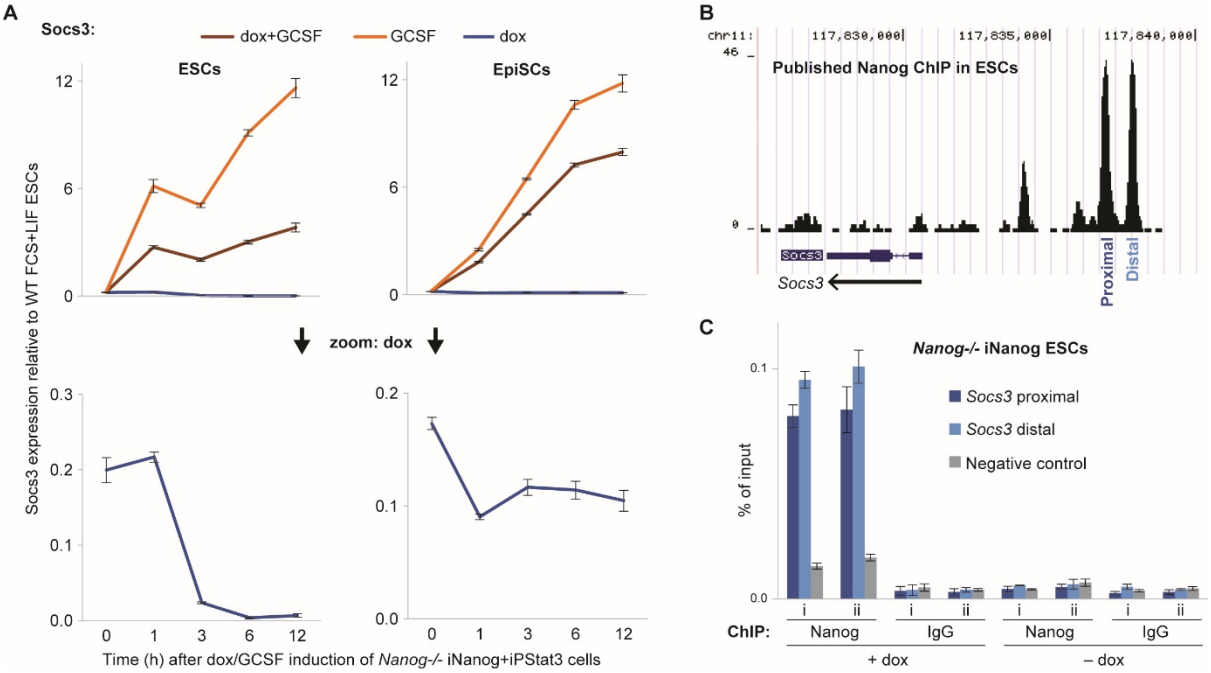


Figure 3.2.3: Nanog binds the Socs3 promoter and represses its transcription

(A) RT-qPCR analysis of Socs3 expression in *Nanog-/-* iNanog+iPStat3 ESCs or EpiSCs following induction with dox and/or GCSF. ESCs were cultured in 2i without LIF, and EpiSCs in FA. Mean expression is shown relative to Gapdh and normalised to wild-type FCS+LIF ESC level=1, ± SD (n=3). Lower panels: zoomed in on the respective dox inductions.
(B) Published ESC ChIP-seq data (Marson et al., 2008) reveal that Nanog binds the upstream regulatory region of the *Socs3* gene. Arrow indicates the direction of Socs3 transcription.
(C) ChIP analysis of Nanog binding at the regulatory region of *Socs3*. Proximal and distal Nanog binding sites are as indicated. Nanog expression was dox-induced in *Nanog-/-* iNanog ESCs in 2i culture and samples were harvested at 0 and 3 hours. Data shown are the mean of 3 technical replicates from each of 2 experiments (i and ii). Error bars indicate ± SD. This ChIP was performed in collaboration with Aliaksandra Radzisheuskaya.

3.2.2 Nanog and PStat3 synergistically upregulate Klf4 and novel *Klf4* enhancer RNA

To further explore the newfound mechanistic link between Nanog and LIF/Stat3 signalling, I investigated the effect of Nanog on expression of LIF/Stat3 targets in ESCs. In the steady-state presence of LIF, I found strong positive correlation between levels of Nanog and Klf4, a canonical PStat3 target (Bourillot et al., 2009; Niwa et al., 2009; van Oosten et al., 2012) (CB Figures 1A, S1B). Interestingly, in response to LIF stimulation, Klf4 upregulation required both Nanog and Stat3 to be present, while Nanog overexpression cooperated with LIF to substantially increase the rate and levels of Klf4 induction (Figure 3.2.4A). Oct4 expression confirmed maintenance of pluripotency (Figure 3.2.4B).

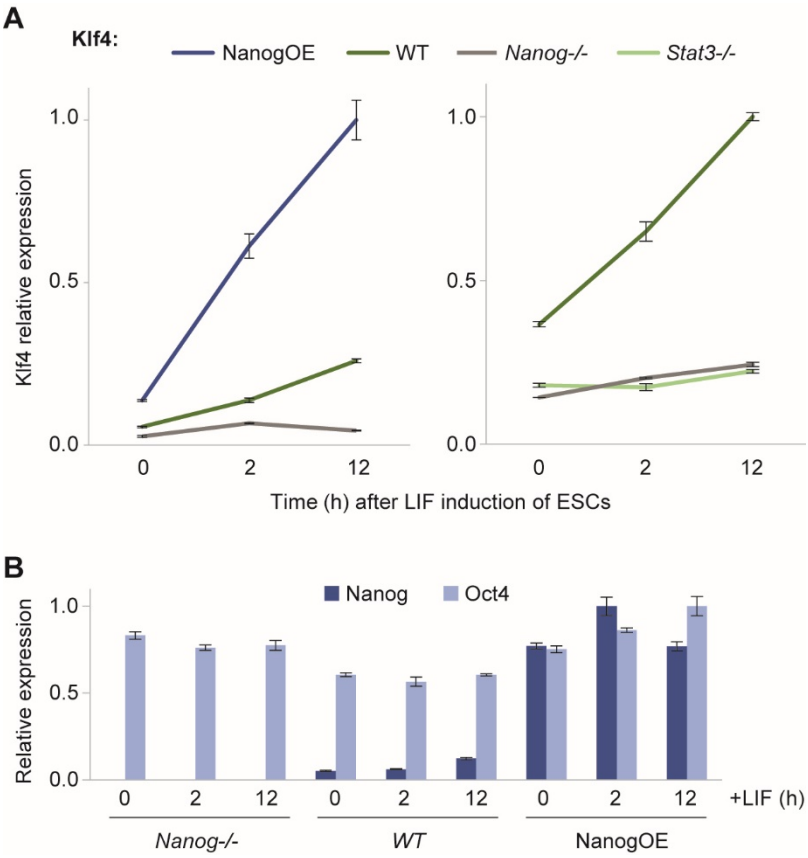


Figure 3.2.4: Nanog modulates the response of Klf4 to LIF

(A–B) RT-qPCR analysis of gene expression, relative to Gapdh and normalised to the highest value. Mean expressions are shown, ± SD (n=3).
(A) Klf4 expression in constitutively Nanog-overexpressing (NanogOE), wild-type (WT), *Nanog-/-* and *Stat3-/-* ESCs in response to LIF stimulation for 0, 2 and 12 hours. Left: LIF was withdrawn from FCS culture for 36 hours prior to re-addition. Selection for pluripotent cells was maintained throughout. Right: LIF was added after at least 7 days in 2i without LIF, because *Stat3-/-* ESCs cannot be cultured in FCS.
(B) Oct4 expression confirms undifferentiated status. Nanog is present as appropriate.

Since Nanog and PStat3 both bind the *Klf4* enhancer (Figure 3.2.5A), it is possible that they regulate Klf4 transcription directly. However, the *Klf4* enhancer lies around 60 kb downstream of the *Klf4* transcription start site and Nanog/PStat3 do not bind the *Klf4* promoter. This prompted two hypotheses, which are not mutually exclusive: (1) formation of a classic enhancer-promoter loop, and/or (2) cis-action of non-coding enhancer RNA. Interestingly, the *Klf4* enhancer was recently identified as an archetypal 'super-enhancer' (Whyte et al., 2013), bound by high levels of the transcriptional coactivator Mediator (Figure 5.2.5A) that may assist in higher-order chromatin organisation. Consistent with hypothesis (1), published Hi-C data indicated a high frequency of physical interaction between *Klf4* enhancer and promoter regions (Figure 3.2.5B) (Dixon et al., 2012).

To test hypothesis (2), I explored the molecular characteristics of the *Klf4* enhancer utilising publicly available ChIPseq and RNAseq data (Figure 3.2.5). RNA PolII binding and H3K27ac co-enrichment suggested active transcription from this enhancer in ESCs, over a 20 kb region (chromosome 4: 55475000–55495000 on assembly mm9). Furthermore, a 'K4-K36 domain' was predicted here, which, in the absence of a known protein-coding gene, is thought to demarcate PolII-transcribed non-coding RNA (Guttman et al., 2009). Indeed, published RNAseq (Martello et al., 2013) indicated bidirectional transcription from the *Klf4* enhancer in ESCs (Figure 3.2.5).

I experimentally confirmed the existence of novel RNA expressed from the *Klf4* enhancer, which I now term *K4eRNA*, by conducting RT-PCR on ESC cDNA template (Figure 3.2.5C). The cDNA was synthesised using random hexamer primers, to avoid polyA+ transcript bias. Genomic DNA was used as a positive control for PCR primers, and cDNA preparations without reverse transcriptase provided a negative control to ensure that RT-PCR signal was solely attributable to RNA, not gDNA contamination. Transcript was detected from mm9 4:55470000–55495000, with a gap from mm9 4:55480000–55484000. This was consistent with RNAseq and PolII ChIPseq data, although the meaning of the gap remains unclear.

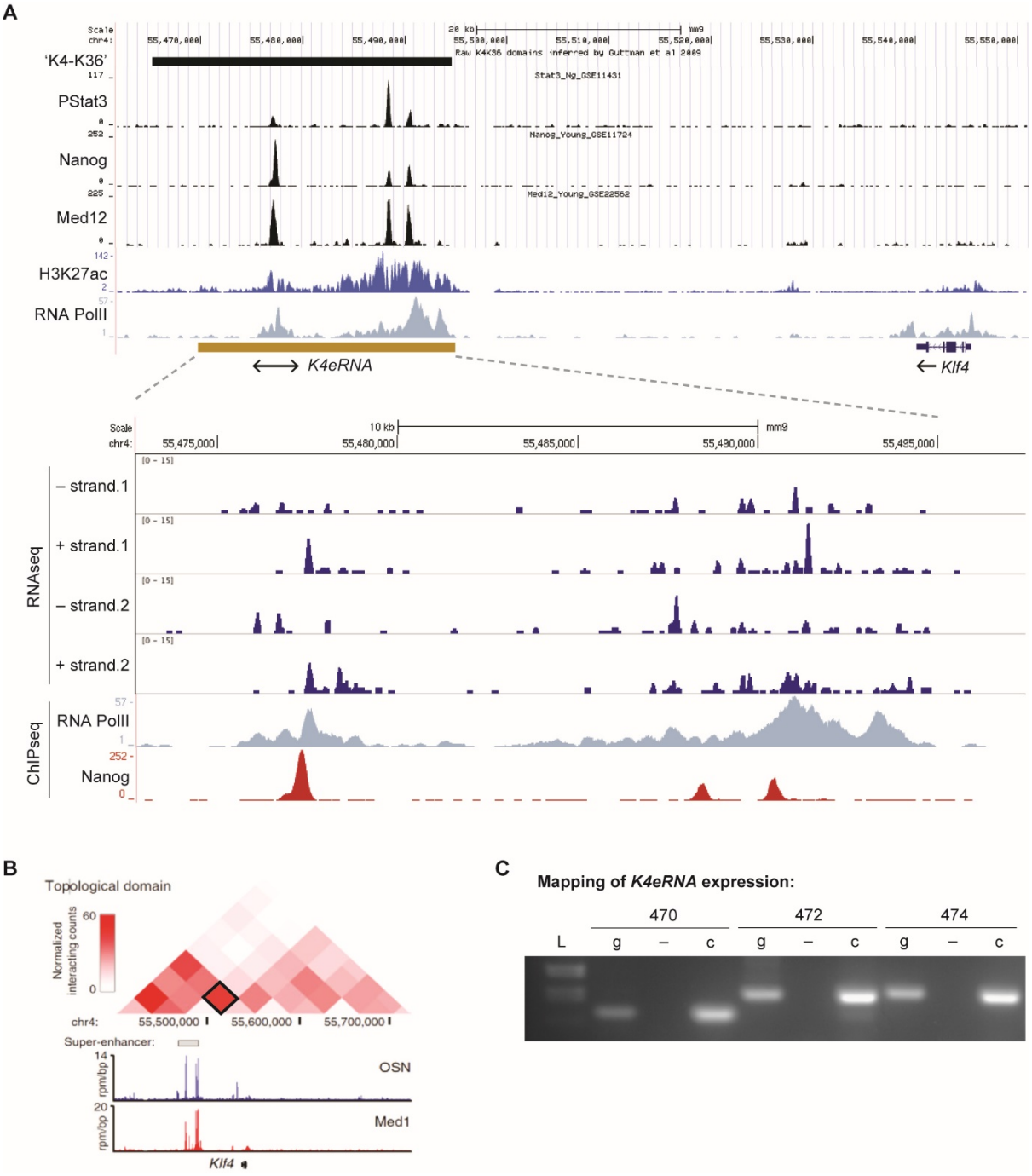


Figure 3.2.5: Identification of novel enhancer RNA at the *Klf4* locus

(A) The *Klf4* locus is shown on assembly mm9, together with publicly available ESC ChIPseq and RNAseq data. Novel enhancer RNA (*K4eRNA*) is indicated by the golden box.
(B) Adapted from Whyte *et al,* 2013, based on data from Dixon *et al*, 2012. Hi-C interaction frequency is displayed above the *Klf4* locus, and shows that the *Klf4* enhancer and promoter regions interact at high frequency (boxed-square) in ESCs. OSN: Oct4/Sox2/Nanog.
(C) Transcription from the *Klf4* enhancer was confirmed by RT-PCR. ESC cDNA (c), genomic DNA (g), and cDNA preparations without reverse transcriptase (–) were used as templates for RT-PCR. Primers were designed to amplify 60–150 bp regions every 1–2 kb from mm9 4: 55470000–55500000. Expression was detected from mm9 4:55470000–55495000, with a gap from mm9 4:55480000–55484000. Representative expressed regions are shown, centred on mm9 4:55470000, 55472000 and 55474000. L: molecular weight ladder.

If *K4eRNA* is responsible for signal transduction between the *Klf4* enhancer and promoter, then its expression should show positive or negative correlation with Klf4, depending on whether it is an activating or repressive regulator. RT-qPCR analysis of *K4eRNA* expression revealed that it positively correlated with Klf4, responding to Nanog and LIF/Stat3 in the same manner (Figure 3.2.6A). Therefore, I hypothesize that *K4eRNA* is a *cis*-activator of Klf4 transcription, since Nanog and PStat3 bind only to the *Klf4* enhancer but not promoter.

The correlation between Nanog and *K4eRNA*/Klf4 was abrogated in the absence of PStat3 in *Stat3+/+* ESCs without LIF and in *Stat3-/-* ESCs (Figure 3.2.6B), demonstrating that Nanog-driven *K4eRNA*/Klf4 upregulation was PStat3-dependent. However, the effect of Nanog on *K4eRNA*/Klf4 transcription cannot be solely attributable to Nanog-mediated PStat3 elevation, since Stat3 hyperactivation in the absence of Nanog could not rescue *K4eRNA*/Klf4 expression (Figure 3.2.6B). Together, this indicates co-dependent action of Nanog and PStat3 at the *Klf4* enhancer in ESCs.

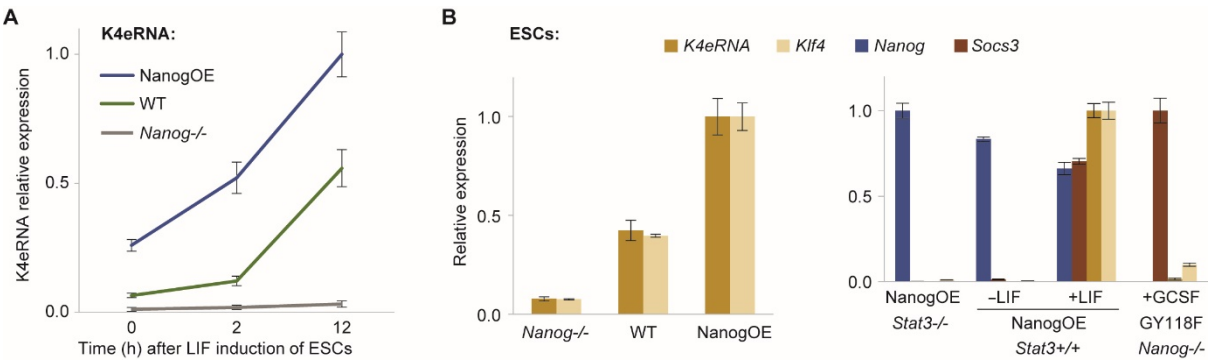


Figure 3.2.6: *K4eRNA* responds to Nanog and PStat3 in the same manner as Klf4

(A–B) RT-qPCR analysis of gene expression, relative to Gapdh and normalised to the highest value. Mean expressions are shown, ± SD (n=3).
(A) *K4eRNA* expression in *Nanog-/-*, wild-type (WT) and constitutively Nanog-overexpressing (NanogOE) ESCs in response to LIF stimulation. LIF was withdrawn from FCS culture for 36 hours prior to re-addition. Selection for pluripotent cells was maintained throughout.
(B) Left: Klf4 and *K4eRNA* expression in *Nanog-/-*, wild-type (WT) and constitutively Nanog-overexpressing (NanogOE) ESCs cultured in steady-state FCS+LIF. Right: Klf4 and *K4eRNA* expression in NanogOE *Stat3-/-*, NanogOE *Stat3*+/+, and Stat3-activated (GY118F+GCSF) *Nanog-/-* ESCs in basic conditions ±LIF. Socs3 expression indicates Stat3 activation as appropriate.

The relationship between Nanog, PStat3 and *K4eRNA*/Klf4 expression was further dissected using *Nanog-/-* iNanog+iPStat3 ESCs in 2i without LIF. Together, dox-induction of Nanog expression and GCSF-stimulation of Stat3 activation elicited *K4eRNA*/Klf4 upregulation in a synergistic manner compared to induction of either factor alone (Figure 3.2.7). This synergistic action of Nanog and PStat3 is specific to *K4eRNA*/Klf4: other pluripotency factors did not respond in this manner, including Nanog-target Esrrb (Festuccia et al., 2012) and PStat3-target Klf5 (Bourillot et al., 2009; Niwa et al., 2009) (Figure 3.2.7).

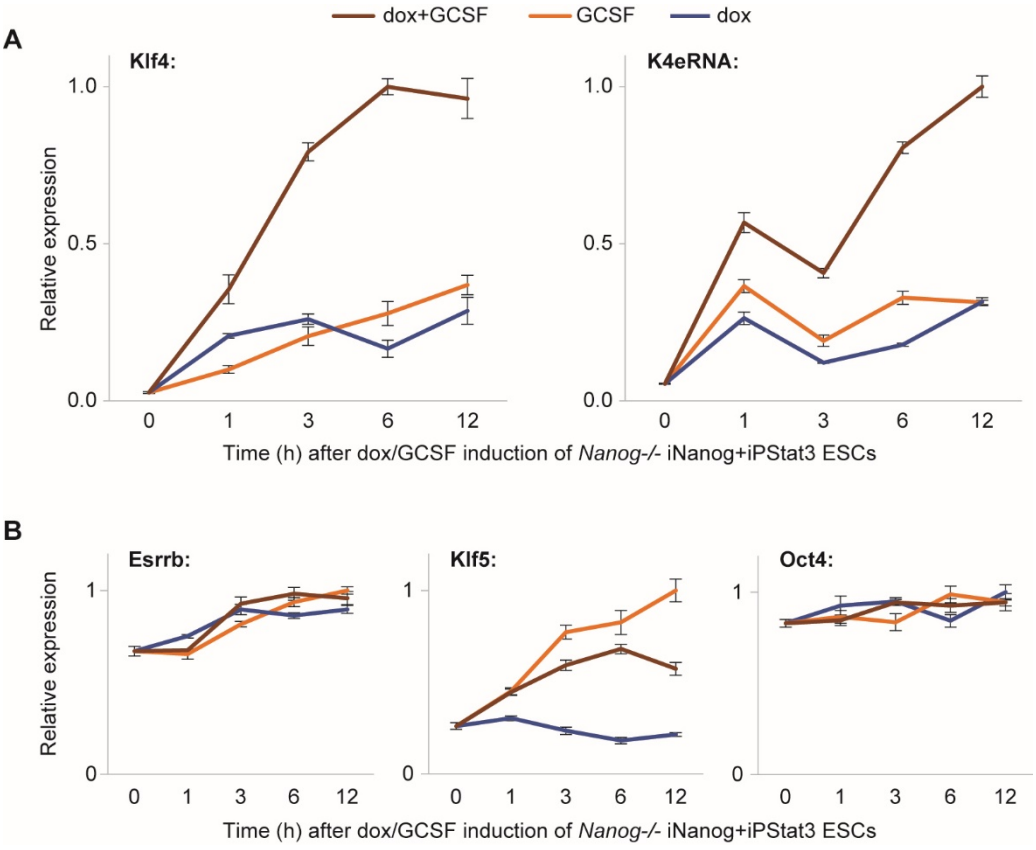


Figure 3.2.7: Nanog and PStat3 synergistically upregulate *K4eRNA* and Klf4

RT-qPCR analysis of mean gene expression, relative to Gapdh and normalised to the highest value, ± SD (n=3). Expression of Klf4 and *K4eRNA* **(A)** or Esrrb, Klf5, and Oct4 **(B)** are shown in *Nanog-/-* iNanog+iPStat3 ESCs following induction with dox and/or GCSF in 2i without LIF.

3.2.3 Nanog and PStat3 induce rapid and efficient reactivation of naïve genes

Having mechanistically linked Nanog and PStat3 in ESCs where the pluripotency network is fully operational, I turned to EpiSCs to study the role of these connected mechanisms during naïve pluripotency establishment. This necessitated generation of *Nanog-/-* background EpiSCs (Figure 3.2.8) to eliminate confounding endogenous Nanog expression.

NanogF/F (Chambers et al., 2003) and *Nanog+/-* (Mitsui et al., 2003) mice were crossed, and EpiSCs were derived from resultant *Nanog-/F* and *Nanog+/F* post-implantation embryos in collaboration with Jenny Nichols. I then treated *Nanog-/F* EpiSCs with TAT-CRE, to excise *LoxP*-flanked Nanog exons and bring *GFP* under control of the *Nanog* promoter. This resulted in *Nanog-/-* EpiSCs with *βgeo* and *GFP* expressions driven by endogenous *Nanog* promoters (Figure 3.2.8A–B). Retention of EpiSC identity was confirmed by RT-qPCR and morphology (Figure 3.2.8C–D).

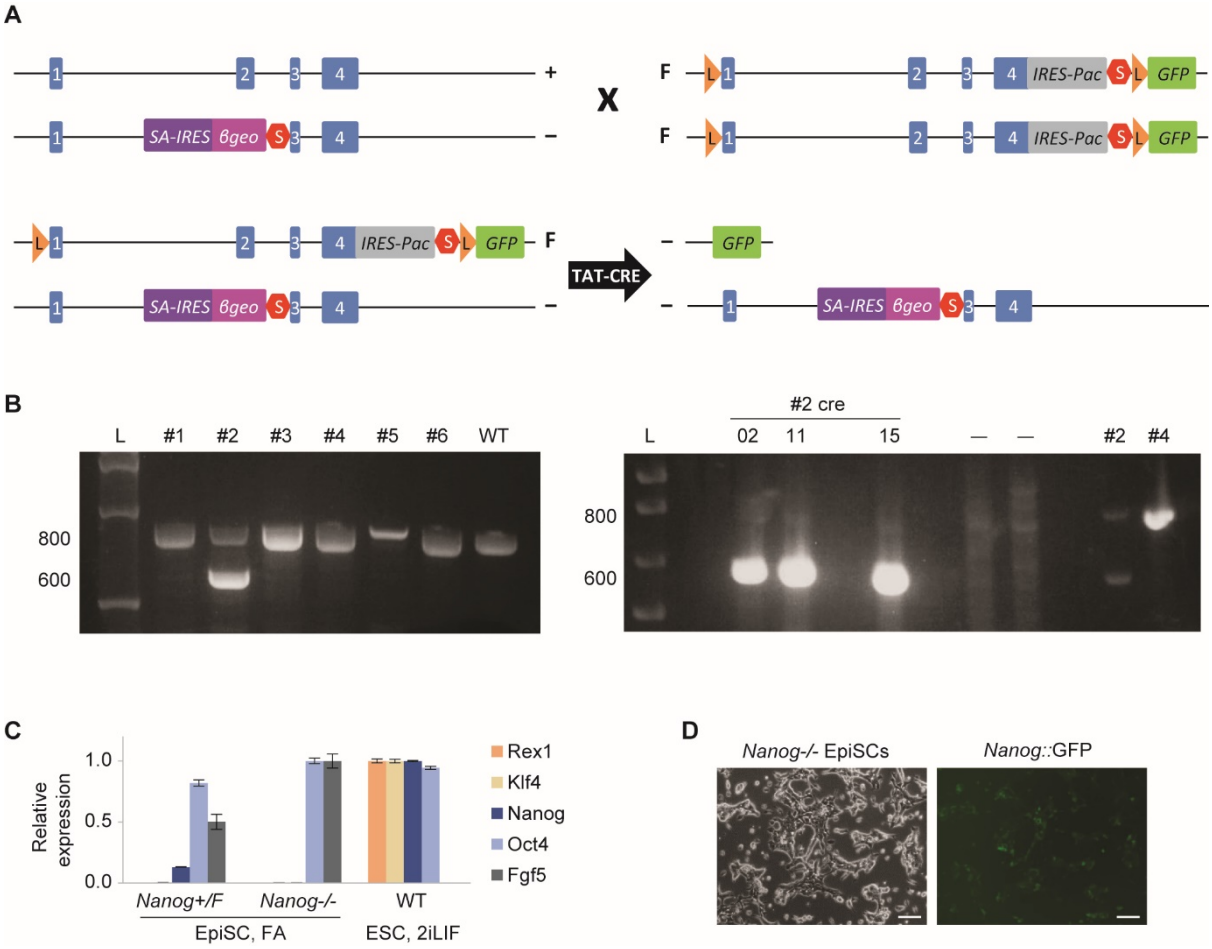


Figure 3.2.8: Generation of *Nanog-/-* embryo-derived EpiSCs

(A) Schematic of the strategy used to generate *Nanog-/-* EpiSCs, not to scale. All alleles are under control of the endogenous *Nanog* promoter, and the *Nanog* 5' UTR is unchanged. Key: floxed-*Nanog* (*F*); wild-type *Nanog* (+); deleted *Nanog* (-); endogenous *Nanog* exons (*1,2,3,4*); splice-acceptor (*SA*); internal ribosome entry site (*IRES*); geneticin resistance (*βgeo*); stop codon (*S*); *LoxP* site (*L*); puromycin resistance (*Pac*). *GFP* coding sequence is brought under control of the endogenous *Nanog* promoter after floxing.
(B) Genotyping PCR was conducted with 3 primers, 1 of which is common to *F*, + and *βgeo* alleles. 1 primer recognizes both *F* and + alleles, to give an 800 bp product. 1 primer anneals to the *βgeo* allele, yielding a 600 bp product. L: molecular weight ladder.
Left: Genotyping of embryo-derived EpiSCs revealed that #2 was *Nanog-/F*. Right: *Nanog-/-* clonal EpiSC lines (#2cre02, 11, and 15) were obtained after treating #2 *Nanog-/F* EpiSCs with TAT-CRE. *Nanog-/-* ESC samples containing different knock-out alleles were included as negative controls (–), while #2 *Nanog-/F* and #4 *Nanog+/F* EpiSCs provided positive controls. *Nanog-/-* #2cre15 EpiSCs were used in all subsequent experiments.
(C) RT-qPCR analysis of *Nanog-/-* and *Nanog+/F* EpiSCs derived from *Nanog-/F* and *Nanog+/F* littermate embryos, cultured in FA. Mean gene expression was measured relative to Gapdh and normalised to the highest value, ± SD (n=3). Lack of Nanog expression in *Nanog-/-* EpiSCs confirmed null genotype, while expression of Oct4 and Fgf5 but not Rex1 or Klf4 confirmed EpiSC identity.
(D) Representative phase and *Nanog::*GFP images of the *Nanog-/-* EpiSCs derived from *Nanog-/F* embryos, in FA conditions. Scale bars: 100 μm.

The aforementioned iNanog+iPStat3 system provides a powerful platform for the quantitative dissection of Nanog and PStat3 mechanisms, since they can be induced separately and in combination within a single cell line. Therefore, I introduced these transgenes to *Nanog-/-* EpiSCs and performed timecourse inductions. In the first instance, I maintained EpiSC FA culture conditions throughout so that responses could be ascribed exclusively to transgene induction.

Strikingly, Nanog and PStat3 co-dependently reactivated Klf4 and *K4eRNA* in this distinct cellular and environmental context (Figure 3.2.9). This demonstrated that their effect on Klf4 was not an ESC-specific phenomenon, and may be of functional relevance for Nanog/PStat3-driven reprogramming.

Known Nanog-target Esrrb (Festuccia et al., 2012) responded to dox-induction of Nanog expression in EpiSCs. Similarly, known PStat3 targets Klf5 (Bourillot et al., 2009; Niwa et al., 2009) and Tfcp2l1 (Martello et al., 2013; Ye et al., 2013) were upregulated following GCSF-induction of Stat3 activation (Figure 3.2.9). However, the synergistic response of Klf4 to Nanog and PStat3 remained unique.

Although Klf2 is not considered a PStat3 target in ESCs (Hall et al., 2009), I found that it was upregulated within 1 hour in response to Stat3 activation in EpiSCs whilst maintaining FA conditions (Figure 3.2.9). This suggests that Klf2 is a target of PStat3 in EpiSCs, raising the possibility of different network topologies during establishment vs maintenance of naïve pluripotency. Unlike Klf4, Klf2 induction in response to PStat3 was dampened rather than augmented by co-induction of Nanog. Another curious case of opposing action between PStat3 and Nanog was proffered by Sox2, a core pluripotency factor expressed in both primed and naïve pluripotent states. Sox2 expression was transiently boosted by PStat3 but repressed by Nanog (Figure 3.2.9). These contradictory effects appeared to balance each other in the scenario of co-induction – please see Section 8.2.2 for related discussion.

In total, Nanog and PStat3 rapidly reactivated many key components of the naïve pluripotency network to near ESC-level, shedding light on their ability to drive fast and efficient reprogramming. This is even more remarkable when taking into account that the assay used EpiSC FA culture conditions instead of conditions promoting reprogramming or ESC self-renewal.

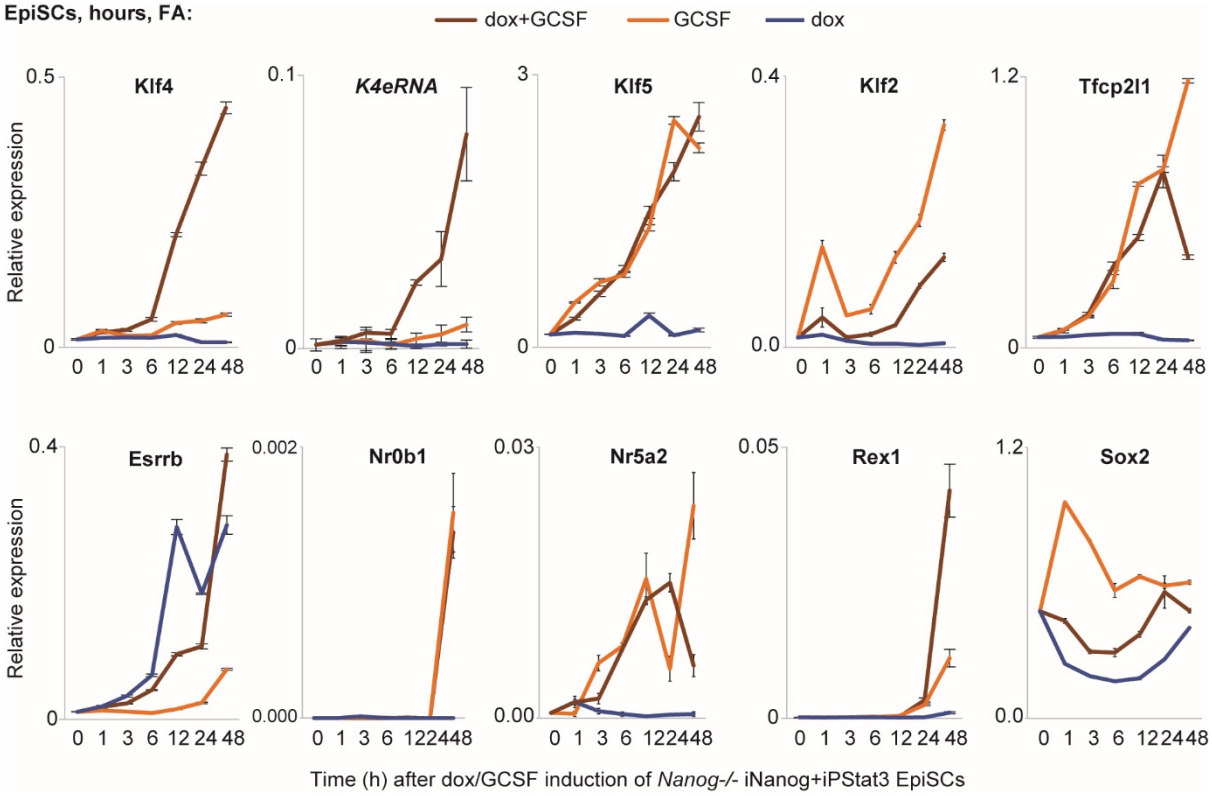


Figure 3.2.9: Nanog and PStat3 induce rapid reactivation of naïve genes in EpiSCs

RT-qPCR analysis of *Nanog-/-* iNanog+iPStat3 EpiSCs following induction with dox/GCSF in FA. Mean expression was measured relative to Gapdh and normalised to FCS+LIF ESC-level=1, ± SD (n=3).

3.2.4 Combined PStat3 and Klf4 bypass Nanog in reprogramming

Although Nanog is dispensable for pluripotency maintenance (Chambers et al., 2007), it is required for establishment of the pluripotent epiblast during preimplantation embryonic development (Silva et al., 2009). Correspondingly, Nanog is essential for naïve pluripotency establishment during conventional *in vitro* reprogramming experiments (Silva et al., 2009). Rescue of *Nanog-/-* reprogramming thus provides a means of functionally testing proposed downstream mechanisms of Nanog.

So far, I have described two new Nanog mechanisms: PStat3 elevation by Socs3 repression, and Klf4 upregulation in cooperation with PStat3. Therefore, I investigated the ability of PStat3 and Klf4 to rescue reprogramming of *Nanog-/-* EpiSCs. I also tested Esrrb, since it has previously been reported as a Nanog downstream target able to bypass Nanog in reprogramming of ESC-derived EpiSCs (Festuccia et al., 2012). Expression of *GFP* and *βgeo* under the control of endogenous Nanog promoters provided visual and selective reporters in

our system, since EpiSCs cannot survive long-term in the reprogramming culture conditions (Figure 3.2.10A).

Forced expression of Nanog, Esrrb, Klf4 or PStat3 individually can drive reprogramming of *Nanog+/+* EpiSCs in 2iLIF conditions (Festuccia et al., 2012; Guo et al., 2009; Silva et al., 2009; Yang et al., 2010). I verified the ability of *Nanog-/-* EpiSCs to generate iPSCs in 2iLIF when rescued by Nanog expression, and confirmed that Stat3 activation in conjunction with Nanog expression drove rapid reprogramming at high efficiency (Figure 3.2.10B). However, in the absence of Nanog, individual and combined expression of Esrrb, Klf4 and PStat3 could not rescue reprogramming in 2iLIF (Figure 3.2.10B).

Since Nanog is a reported target of PD03 (Lanner et al., 2010; Silva et al., 2009), I hypothesised that PD03 may be detrimental in the absence of Nanog and thus tested PD03-free reprogramming conditions. I tried to rescue *Nanog-/-* reprogramming with individual and combined expression of Esrrb, Klf4 and PStat3 in Chiron (CH)+LIF (Figure 3.2.10C) and in FCS+LIF. To my surprise, I found that Esrrb overexpression was unable to drive *Nanog-/-* EpiSC reprogramming in any condition. I speculate that this is due to differences in reprogramming propensity between these EpiSCs derived from *Nanog-/F* embryos, and ESC-derived secondary EpiSC systems (Festuccia et al., 2012). Individually, Klf4 and PStat3 also failed to yield iPSCs. Although they initially generated GFP+ colonies in CH+LIF (Figure 3.2.10C), these lacked the capacity to self-renew after passaging into 2iLIF, suggesting that reprogramming was incomplete.

In combination, PStat3+Esrrb and PStat3+Klf4 yielded *Nanog-/-* GFP+ iPSCs in CH+LIF conditions (Figure 3.2.10C). Although *Nanog-/-* iPSCs could not be established in 2iLIF or PD03+LIF, once naïve pluripotency was established in CH+LIF, *Nanog-/-* iPSCs could be passaged indefinitely in 2iLIF, consistent with known discrepancies between Nanog requirement in pluripotency establishment vs maintenance. After passaging in 2iLIF, gene expression profile and observed chimeric competence of *Nanog-/-* iPSCs formally demonstrated their reacquisition of a naïve pluripotent program (Figure 3.2.10D–E).

Together, this highlights PStat3 activation as a key functional mechanism acting downstream of Nanog, which, in conjunction with overexpression of either Nanog-target Klf4 or Esrrb, can efficiently rescue *Nanog-/-* reprogramming. It is of interest to note that Klf4 but not Nanog overexpression can drive EpiSC reprogramming in the presence of JAK inhibitor (Yang et al., 2010), further supporting the placement of Nanog upstream and Klf4 downstream of Stat3 activation during reprogramming.

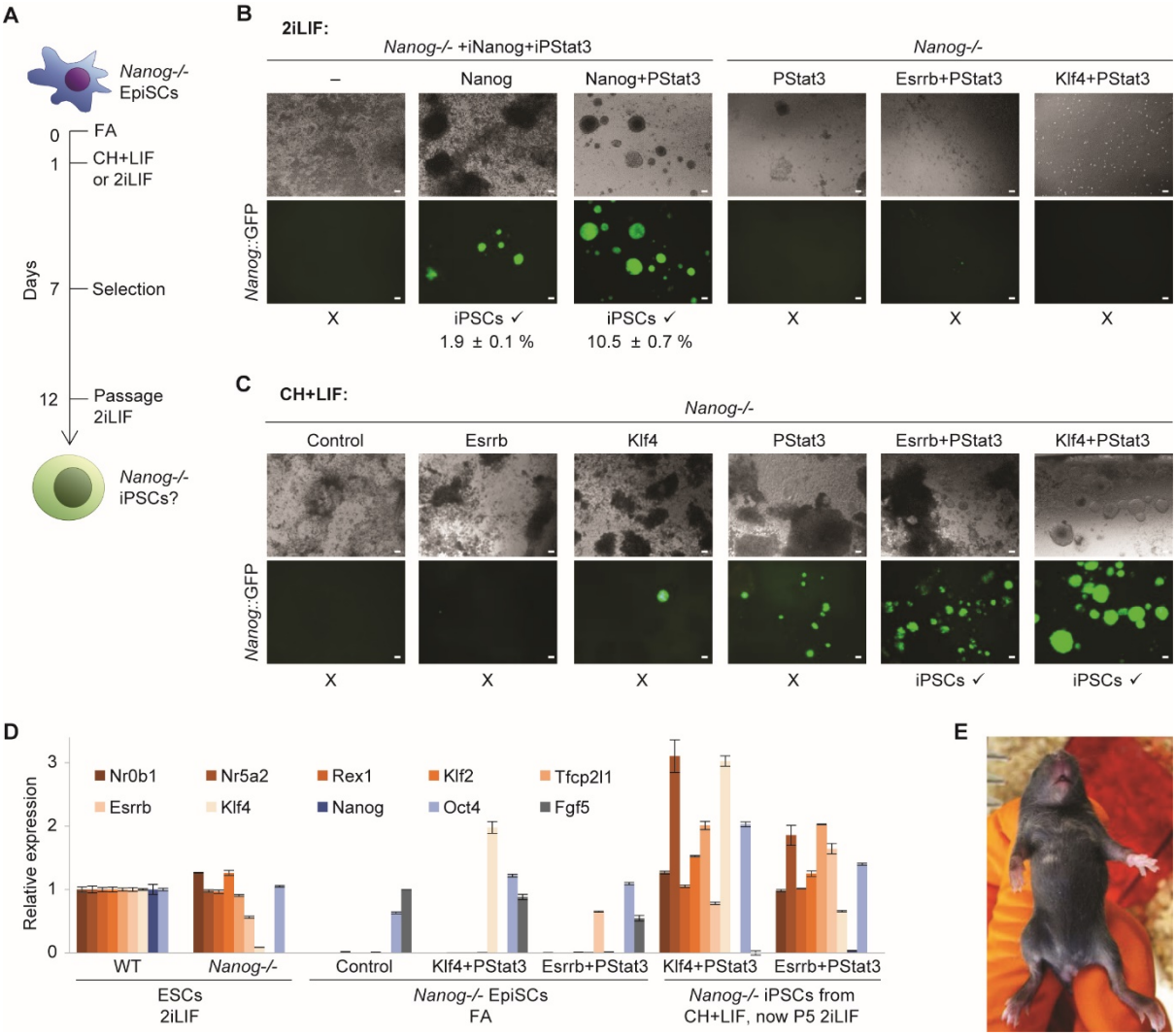


Figure 3.2.10: Combined PStat3 and Klf4 bypass Nanog in reprogramming

(A) Schematic depicting the protocol for reprogramming of *Nanog-/-* EpiSCs containing *βgeo* and *GFP* reporters under the endogenous *Nanog* promoters.
(B) Representative images taken on day 12 of reprogramming in 2iLIF. The mean number of GFP+ iPSC colonies is indicated as % of cells initially plated, ± SD (n=3). Scale bars: 100 µm.
(C) Representative images taken on day 12 of reprogramming in CH+LIF. Quantification of reprogramming efficiency by GFP+ colony counting was inappropriate in this case, since CH+LIF is a permissive medium compared to 2iLIF and yielded some GFP+ colonies that were not true iPSCs. After passaging in 2iLIF, clean iPSC lines were obtained as indicated. Scale bars: 100 µm.
(D) RT-qPCR analysis of gene expression in CH+LIF-derived *Nanog-/-* iPSCs after passaging in 2iLIF, compared to parental EpiSCs in FA and wild-type and *Nanog-/-* ESCs in 2iLIF. Expression patterns of Nanog, Esrrb and Klf4 verified the genotypes. Full reprogramming to naïve iPSCs was confirmed by reactivation of naïve genes, maintenance of Oct4, and repression of Fgf5 expression. Mean gene expression ± SD (n=3) is shown relative to Gapdh and normalised to wild-type ESC level, except Fgf5 which is normalised to control EpiSCs.
(E) Chimeric mouse obtained after injecting PStat3+Klf4 EpiSC-derived *Nanog-/-* iPSCs into C57/BL6 blastocysts. Agouti coat colour indicates chimeric contribution (MF1/129 background). The mouse is scruffed with an orange glove.

Nanog is dispensable for the initial formation of reprogramming intermediates (pre-iPSCs) from somatic cells, but is essential for pre-iPSCs to transit to naïve pluripotency in 2i+LIF (Silva et al., 2009). In an exciting parallel to my findings in EpiSCs, collaborators demonstrated that *Nanog-/-* pre-iPSCs could also reprogram in alternative conditions (CH+LIF or KSR+LIF), albeit with low speed and efficiency (CB Figure S4C–D). Microarray analysis of our combined samples revealed that pre-iPSC-derived and EpiSC-derived *Nanog-/-* iPSCs clustered closely with both wild-type and *Nanog-/-* ESCs, demonstrating reprogramming to a naïve pluripotent identity (Figure 3.2.11). It should be noted that retroviral Klf4 and exogenous LIF provided the reprogramming impetus for these pre-iPSCs, again implicating PStat3 and Klf4 in the rescue of *Nanog-/-* reprogramming.

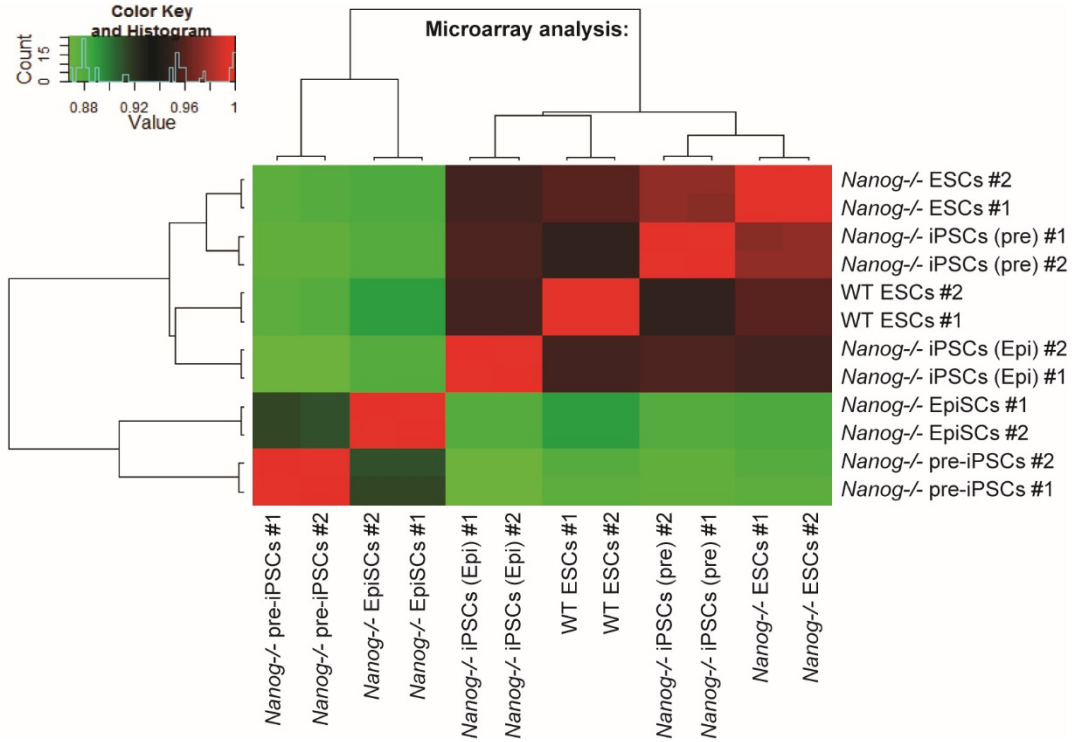


Figure 3.2.11: Acquisition of naïve gene expression signature in *Nanog-/-* iPSCs

Correlation heatmap of gene expression between: wild-type (WT) ESCs, *Nanog-/-* ESCs, pre-iPSC-derived *Nanog-/-* iPSCs, PStat3+Klf4 EpiSC-derived *Nanog-/-* iPSCs all in 2iLIF; parental *Nanog-/-* pre-iPSCs in FCS+LIF; parental *Nanog-/-* EpiSCs in FA. 2 technical replicates were included for each cell line. Pre-iPSC and derived iPSC samples were generated by collaborators Aliaksandra Radzisheuskaya and Anzy Miller.

3.3 Discussion

3.3.1 Modulation of LIF signal transduction by Nanog

Two mechanisms of signal/TF interplay are exemplified by the work of Chapter 3: (1) modulation of extrinsic signal sensitivity by a nuclear TF, and (2) cooperation between the nuclear TF and the activated TF downstream of the extrinsic signal transduction cascade. This highlights the importance of considering the intrinsic properties of a cell when computing the efficacy and output of signal exposure. A signal does not simply give the cell an instruction, but rather is interpreted as a function of the internal network state, even when the basic transduction machinery is already present.

In this particular case, I demonstrated that Nanog enhanced transduction of extracellular LIF signal by transcriptionally downregulating the Socs3 repressor, and additionally cooperated with resultant active PStat3 to elicit synergistic and co-dependent upregulation of Klf4. This mechanistically connects Nanog and PStat3, two important factors for the establishment and maintenance of naïve pluripotency (Figure 3.1.1). Their convergence on Klf4 lends greater biological significance to the inclusion of Klf4 in the canonical 'Yamanaka' cocktail for somatic cell reprogramming.

These findings may provide some insight into how Stat3 signalling has different consequences in different contexts. Here, Nanog potentiates the positive input of PStat3 to the naïve pluripotent network. In contexts without Nanog, PStat3 positively regulates specification of other cell types such as primitive endoderm, astrocytes and inflammatory Th17 T cells (Bonni et al., 1997; Egwuagu, 2009; Morgani and Brickman, 2015). In contrast, LIF/Stat3 and Nanog are again both important for PGC specification, a process with many recapitulated features of naïve pluripotency induction (Chambers et al., 2007; Hayashi et al., 2011; Murakami et al., 2016; Ohinata et al., 2009). It would be of interest to test whether Nanog is involved in restricting PStat3 to its pluripotency-related target genes, and whether Nanog also amplifies LIF/Stat3 action during PGC specification.

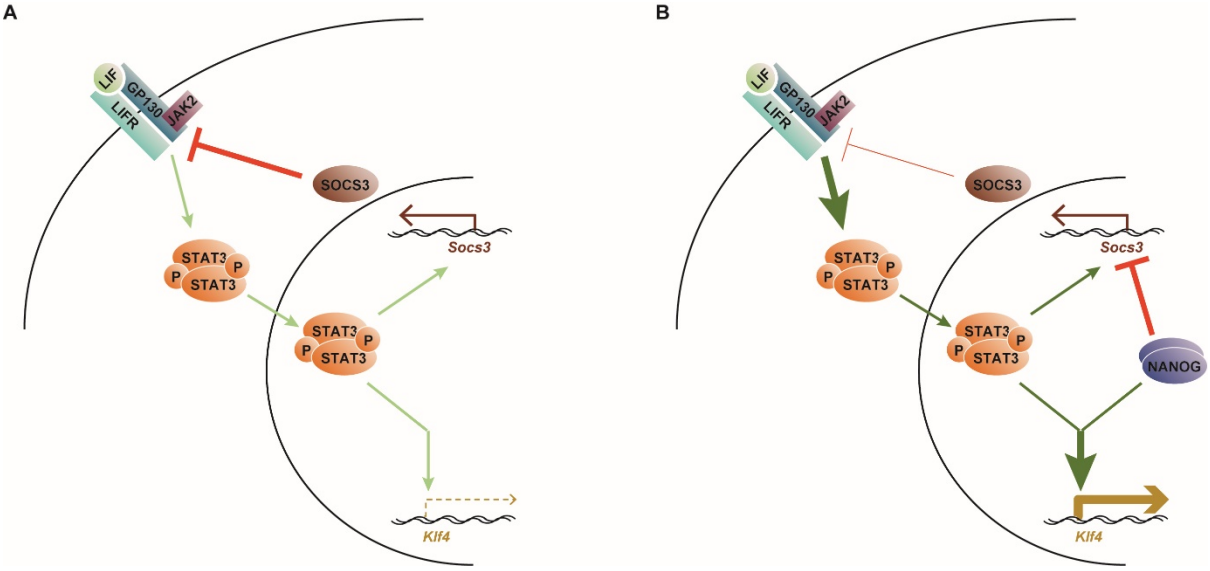


Figure 3.3.1: Newfound role of Nanog in the modulation of LIF/Stat3 signal transduction

Schematics depicting LIF/Stat3 signal transduction to target genes *Socs3* and *Klf4* in the absence **(A)** and presence **(B)** of Nanog. The work of Chapter 3 demonstrated that Nanog represses Socs3 and amplifies LIF signal transduction, resulting in higher levels of active PStat3. Furthermore, I found that Nanog and PStat3 upregulate Klf4 in a co-dependent and synergistic manner. Functionally, PStat3 elevation and Klf4 expression allowed bypass of the requirement for Nanog in reprogramming.

3.3.2 Multifaceted relationship between Nanog and the environment

Nanog thus contributes to the naïve TF network by concurrently inducing the expression of Esrrb (Festuccia et al., 2012), enhancing LIF/Stat3 signal transduction, and inducing Klf4 expression in cooperation with PStat3 (Figure 3.3.1). Ultimately, combinations of these factors allowed *in vitro* bypassing of Nanog for the establishment of a naïve pluripotent cell state (Figure 3.2.10, CB Figure S4C–D). However, the slower kinetics and reduced efficiency of *Nanog-/-* somatic reprogramming implied the existence of further mechanisms by which Nanog operates. These may include additional downstream effectors of reprogramming, and the activities of Nanog partners such as Tet1/2 (Costa et al., 2013).

Intriguingly, 2i conditions abolished *Nanog-/-* reprogramming rescue, whereas usually 2i promotes naïve pluripotency induction (Silva et al., 2008). This condition-specific rescue adds another layer of complexity to TF/signal interplay. PD03 appeared to be the problematic component in the absence of Nanog (Figure 3.2.10). How does loss of a TF flip the impact of a signal from beneficial to detrimental? Nanog is reportedly upregulated in response to PD03 (Lanner et al., 2010; Silva et al., 2009); perhaps this is essential for the positive role of PD03 during reprogramming.

Recently, PD03 has been implicated in both appropriate and inappropriate hypomethylation of naïve ESCs (Choi et al., 2017; Ficz et al., 2013). I wonder whether there is a concerted action of Nanog and PD03 on the epigenome that becomes deleteriously dysregulated in the *Nanog-/-* context. Upon further exploration of the microarray data (Figure 3.2.11), I noticed that several genes pertaining to epigenetic regulation were differentially expressed between wild-type and *Nanog-/-* ESCs in 2iLIF. *Nanog-/-* ESCs exhibited higher expression of DNA methyltransferases, lower expression of methylation-sensitive genes, lower expression of PGC-related genes, and aberrant overexpression of the imprinted H19/Igf2 locus. Preliminary RT-qPCR analysis corroborated the inverse relationship between de novo methyl transferase 3b (Dnmt3b) and methylation-sensitive Dazl expression in response to Nanog modulation. This was condition dependent, evident in 2iLIF but not FCS+LIF.

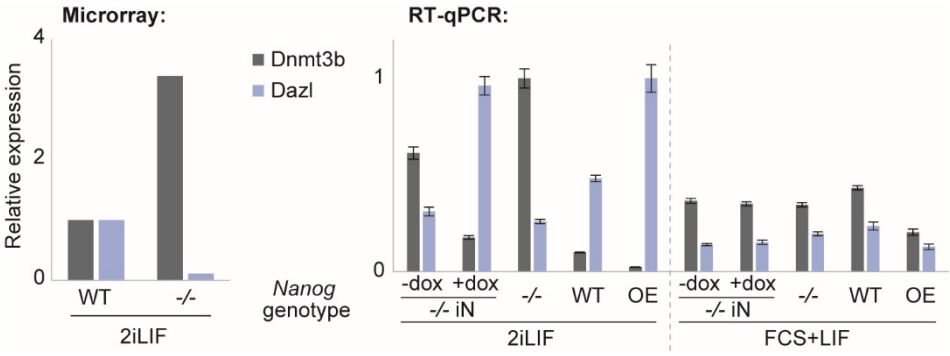


Figure 3.3.2: Condition-dependent relationship between Nanog and epigenetic regulation?

Microarray and RT-qPCR gene expression analysis of Dnmt3b and Dazl in *Nanog-/-*, WT, NanogOE and *Nanog-/-* iNanog ESCs in 2iLIF or FCS+LIF ± dox for 7 days. For microarray data, expression is shown relative to WT ESCs. For RT-qPCR, mean expression was measured relative to Gapdh and normalised to the highest value, ± SD (n=3).

In this epigenetic light, it is interesting to note that bypass of Nanog in somatic reprogramming was enhanced by KSR medium (CB Figure S4C). KSR contains ascorbic acid, a co-factor of dioxygenases including the jumonji histone demethylases and Tets (Blaschke et al., 2013; Esteban and Pei, 2012; Yin et al., 2013). Together, this raises the possibility that Nanog does not simply modulate canonical signal transduction to downstream transcriptional targets, but may also act as a hub to convey information from the environment to the epigenetic landscape.

3.3.3 Klf4-directed synergy between Nanog and PStat3

Induction of Nanog and PStat3 in EpiSCs led to rapid upregulation of many naïve TF network components (Figure 3.2.9). Some of these were attributable to either Nanog or Stat3, as expected for their direct target genes, whereas Klf4 induction resulted from a co-dependent synergy. The relative reprogramming contributions of Klf4 vs other targets of Nanog/PStat3 should be further defined, for example by generation of *Klf4-/-* EpiSCs. Like Nanog and PStat3 themselves, the Klf4-directed synergy may be of greater functional relevance during establishment rather than maintenance of naïve pluripotency (Figure 3.2.6B).

Synergistic upregulation was not observed at other shared targets of Nanog/PStat3, such as Socs3 (Figure 3.2.3), and thus must be specified by genomic sequences or other bound factors at the *Klf4* locus. Since Nanog and PStat3 proteins do not directly bind each other it seems unlikely that they regulate each other's recruitment to the *Klf4* enhancer (Torres and Watt, 2008). Indeed, preliminary ChIP experiments indicated that PStat3 enrichment at the *Klf4* enhancer was not significantly modulated by Nanog expression in ESCs (Graziano Martello, unpublished). Perhaps *Klf4* transcription is regulated by a multicomponent complex, which either cannot form or function positively in the absence of Nanog or PStat3. Meanwhile, comparative study of the gene regulatory sequences at other shared Nanog/PStat3 targets could pinpoint the special features of the *Klf4* locus that cause it to respond in this manner. Such questions are at the core of understanding how differential gene regulation is achieved by a limited set of expressed factors (Section 1.1).

Preliminary evidence suggested that the Nanog/PStat3/Klf4 regulatory relationship may also operate in other contexts of naïve pluripotency establishment, including during *in vivo* epiblast emergence and resetting of conventional primed human ESCs. In agreement with my findings *in vitro*, *Nanog-/-* murine blastocysts failed to activate PStat3 in their ICM (Kurowski *et al*., in preparation). Meanwhile, single-cell RNAseq revealed a reduced strength of positive correlation between Nanog and Klf4 expression in *Stat3-/-* ICMs (Boroviak *et al.*, in preparation). In primed human ESCs, co-induction of Nanog+PStat3 led to synergistic upregulation of both Klf4 and Klf2, respectively indicating conservation and divergence of naïve network induction logic between species (Figure 3.2.9 vs 3.3.3). Further work will be required to unravel the functional relevance of these observations and how they fit within their respective network topologies.

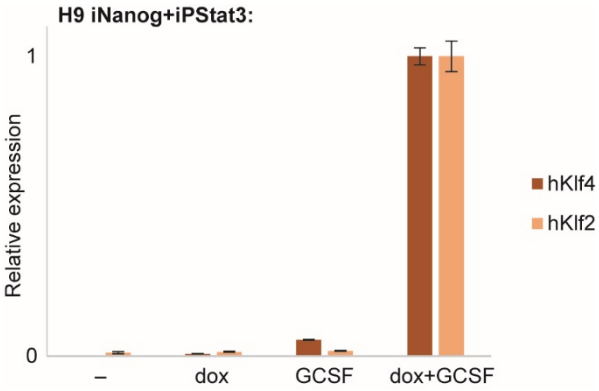


Figure 3.3.3: Synergistic response of Klfs to Nanog and PStat3 in human ESCs

H9 conventional primed human ESCs were transfected with iNanog and GY118F (iPStat3), stably selected, then treated with dox and/or GCSF for 24 hours whilst maintaining primed culture conditions. Mean Klf4 and Klf2 expression were measured by RT-qPCR relative to Gapdh and normalised to the highest value, ± SD (n=3). This was conducted in collaboration with Tim Lohoff.

3.3.4 Role and conservation of the novel *Klf4* enhancer RNA

This work identified novel RNA expressed from the *Klf4* enhancer (*K4eRNA*), with preliminary evidence suggestive of its role as a *cis*-activator of *Klf4* transcription (Section 3.2.2). Unfortunately, attempts to knock-down *K4eRNA* were unsuccessful for technical reasons, thus precluding confirmation of this hypothesis. Studies conducted on bidirectionally transcribed eRNAs in other cell types have found that only one of the two strands is functional, but this does not necessarily correlate with orientation relative to the *cis* target (Lam et al., 2013; Li et al., 2013). Further *K4eRNA* transcript mapping and sequence analysis may assist in the identification of key regions to target on each strand.

As outlined earlier, the *Klf4* enhancer could regulate Klf4 transcription by expression of *cis*-activating *K4eRNA*, or by formation of a classic enhancer-promoter loop via Mediator. In fact, these two hypotheses can be amalgamated: it was recently demonstrated that enhancer-expressed non-coding RNA can activate its target gene in *cis* by directly binding to and recruiting Mediator, modulating the kinase activity of Mediator towards histone H3 serine 10, and promoting Mediator-regulated DNA looping between the enhancer and promoter (Lai et al., 2013). Additionally, it was recently proposed that functional eRNA can induce enhancer-promoter looping in a TF- and cohesin-stabilized manner (Li et al., 2013). It will be of interest to examine whether *K4eRNA* is targeted to the *Klf4* promoter in this manner via Mediator or cohesin, by another protein intermediate, or by direct nucleic acid binding. First, RNA and DNA fluorescence *in situ* hybridisation should be performed to assess *K4eRNA* localisation relative

to the *Klf4* enhancer and promoter. According to Hi-C analysis of three-dimensional higher order genomic organisation, a DNA loop does form between the *Klf4* enhancer and promoter in ESCs (Figure 3.2.5B) (Dixon et al., 2012; Whyte et al., 2013).

More is known about the mechanistic action of repressive than activating non-coding RNAs, such as those involved in X chromosome inactivation and imprinting. It is plausible that, like repressive non-coding RNAs, activating eRNAs recruit epigenetic modifiers or remodellers to their target locus. Around the *Klf4* transcription start site in FCS+LIF ESCs, there is a conspicuous 'poised' bivalent chromatin domain of overlapping repressive H3K27me3 and activating H3K4me3 (publicly available ESC ChIP-seq data). Therefore, additional hypotheses are that *K4eRNA* recruits H3K27me3-demethylases, or H3K4-methylases, to tip the 'poised' domain towards a more active chromatin state. Recruitment of chromatin modifiers could be accomplished by a different region of the K4eRNA to that responsible for targeting to the *Klf4* gene, and may be better conserved on the structural than sequence level. Thus far, very few functional eRNAs have been identified, and what little is known about their mechanisms of action indicates that they do not all act in the same fashion (Mousavi et al., 2013). Therefore, if *K4eRNA* is functional, dissection of the molecular mechanism by which it *cis*-activates *Klf4* transcription could reveal novel gene regulatory mechanisms of broad relevance.

For a preliminary assessment of whether *K4eRNA* is conserved in other species, I aligned the murine *Klf4* enhancer sequence against animal genomes available in 2013, using Ensembl Comparative Genomics tools. Expressed regions of the enhancer sequence are well conserved in the *Klf4* enhancers of all mammals, including the monotreme platypus (Figure 3.3.4). Small conserved regions are found downstream of *Klf4* in bird, reptile and amphibian genomes, but I could not find conserved sequence in fish. In mammals, birds, reptiles and amphibians, *Klf4* is found in a gene desert, with the putative *K4eRNA* sequence far downstream of *Klf4*. Since it is unusually well conserved, gene organisation around the *Klf4* locus is likely to be of functional importance. In contrast, organisational constraints do not apply in fish, where *Klf4* is found in typical gene-rich regions, and other protein-coding genes can be found immediately downstream of *Klf4* (Figure 3.3.4). In the future, it will be of interest to determine the molecular mechanism of *K4eRNA* action, whether it is indeed expressed from these conserved enhancer sequences in other species, and whether eRNA expression is a general feature of 'super-enhancer' functionality both across the murine genome and across species.

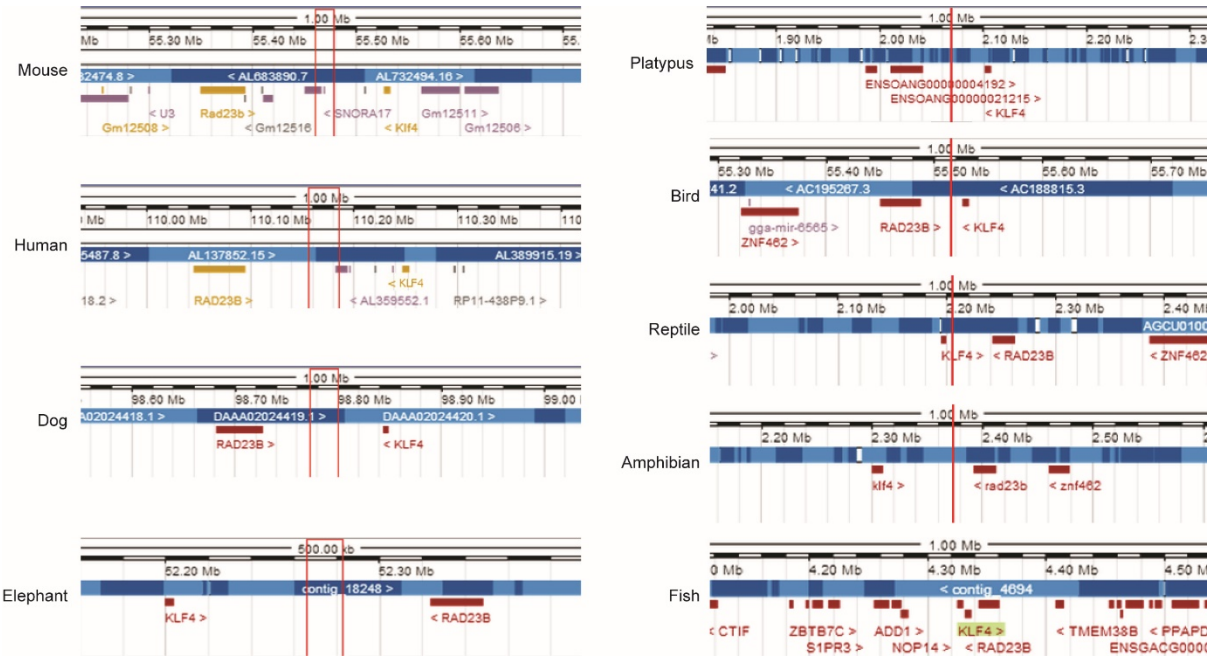


Figure 3.3.4: Conservation of the *K4eRNA* locus

The DNA sequence encoding murine *K4eRNA* was aligned against available animal genomes using Ensembl Comparative Genomics tools on default settings on 01.06.13. Conserved regions are delineated by red boxes. Regions of the murine *K4eRNA* sequence are well conserved in the *Klf4* enhancers of all mammals, including the monotreme platypus. Small conserved regions are found downstream of *Klf4* in available bird, reptile and amphibian genomes, but not in fish: representative examples are shown here (chicken, Chinese soft-shelled turtle, Xenopus and stickleback respectively). Gene organisation around the *Klf4* locus is unusually well conserved in mammals, birds, reptiles and amphibians, but not in fish. Genomes are displayed on the most up-to-date assemblies as of 01.06.13 (mm10 rather than mm9 for mouse). Unfortunately I did not take a note of the default settings of the Ensembl Comparative Genomics tools on 01.06.13 and therefore this Figure should be considered as strictly preliminary and for interest only.

CHAPTER 4:
GENERATION OF A TRACTABLE IDENTITY TRANSITION SYSTEM

4.1 Introduction

As outlined in Chapter 1, a core aim of this thesis is to understand how TFs and signals interact to specify a cellular identity. The results in Chapter 3 provided an example of exogenous signal transduction modulation (LIF/Stat3) by a nuclear TF (Nanog). Importantly, this work also laid the groundwork for the utility of EpiSC reprogramming as an identity transition model for mechanistic study (Section 1.3). In Figure 3.2.9, I showed for the first time that the kinetics of naïve network reactivation in EpiSCs are rapid enough to study direct effects of factor induction on destination program installation, even whilst maintaining EpiSC FA conditions. Therefore, by independently varying factor inductions with signal perturbations, it should be feasible to disentangle the contributions of TFs and the environment to transcriptional rewiring. I test this in Section 4.2.1.

A number of technical barriers remained to the causal connection of molecular events to individual inputs (Section 1.3). Therefore, for the remainder of Chapter 4 I focus on solving these problems and developing appropriate tools to answer my biological questions.

Compared to somatic cell reprogramming, EpiSC reprogramming is relatively efficient and thus better suited to mechanistic dissection of identity transitions. However, even in the best case scenarios, EpiSC reprogramming had so far not been achieved with efficiencies over ~10% (Section 1.2.8). Due to the resultant heterogeneity, averages are confounded and thus cells undergoing the change of interest must be resolved from the bulk in order to study transition mechanism(s) (Figure 4.1.1). With this motivation, I establish and validate methodology to isolate productive EpiSC reprogramming intermediates in Section 4.2.2.

When considering how interplay between TFs and signals elicit a cell identity transition, a fundamental outstanding question is whether there are multiple mechanisms by which to undergo a given identity transition. More detailed introductions to this concept can be found in Chapter 1 and in Section 5.1. Here, in Sections 4.2.3–5, I develop and optimise the tools that will be required to answer this question, i.e. different inducible single-factor reprogramming drivers of high efficiency, whose mechanisms of reprogramming can later be compared.

Figure 4.1.1: Productive intermediates must be isolated from the bulk

Schematic illustrating the heterogeneous and asynchronous nature of reprogramming, necessitating a method by which to isolate productively transitioning intermediates (greens) for mechanistic study.

4.2 Results

4.2.1 Interaction between TFs and signalling environment during reprogramming

In Figure 3.2.9, I assessed the initial transcriptional response of *Nanog-/-* iNanog+iPStat3 EpiSCs following inductions whilst maintaining EpiSC FA culture conditions. Having established this baseline for the responses attributable solely to transgenes, I then extended the analysis to intersect with the contribution of signal variables to naïve network reactivation during reprogramming. First I confirmed the ability of Nanog and PStat3 to dominantly drive reprogramming despite FA conditions (Figure 4.2.1A–C). As expected (van Oosten et al., 2012) reprogramming occurred, but at substantially lower efficiency than in 2iLIF. Nonetheless, resultant iPSCs exhibited naïve expression profile and chimera competence (Figure 4.2.1D–E) thus validating FA+dox+GCSF as a reprogramming condition.

Following Nanog+PStat3 induction with dox+GCSF in either FA or 2iLIF, bulk timecourse expression analysis over 6 days revealed upregulation of naïve genes in both conditions. Some genes initially responded in a similar manner irrespective of environment (Klf4, *K4eRNA*, Klf2, Rex1), suggestive of early regulation by TF drivers rather than the environment. The divergence in these upregulation kinetics after day 3 may reflect the ultimately higher reprogramming efficiency in 2iLIF than FA and/or a later signal sensitivity. In contrast, Tfcp2l1 and Nr5a2 were positively regulated by 2iLIF throughout, whereas Klf5 and Esrrb were damped initially.

The responses of Sox2 and Oct4 were extremely surprising, both undergoing substantial downregulation in 2iLIF compared to FA. Given the higher reprogramming efficiency in 2iLIF (Figure 4.2.1B) these profiles in the bulk averages were counterintuitive. Notably, as core pluripotency factors, Oct4 and Sox2 are expressed in primed EpiSCs as well as naïve iPSCs. Considering this, two hypotheses came to mind: (1) transient suppression of Oct4/Sox2 may be necessary to dismantle the self-renewing primed network and rebuild the naïve network, which supports Oct4/Sox2 in a different configuration and from different enhancer elements, or (2) the Oct4/Sox2 drop following signal switch is a barrier to be overcome in reprogramming.

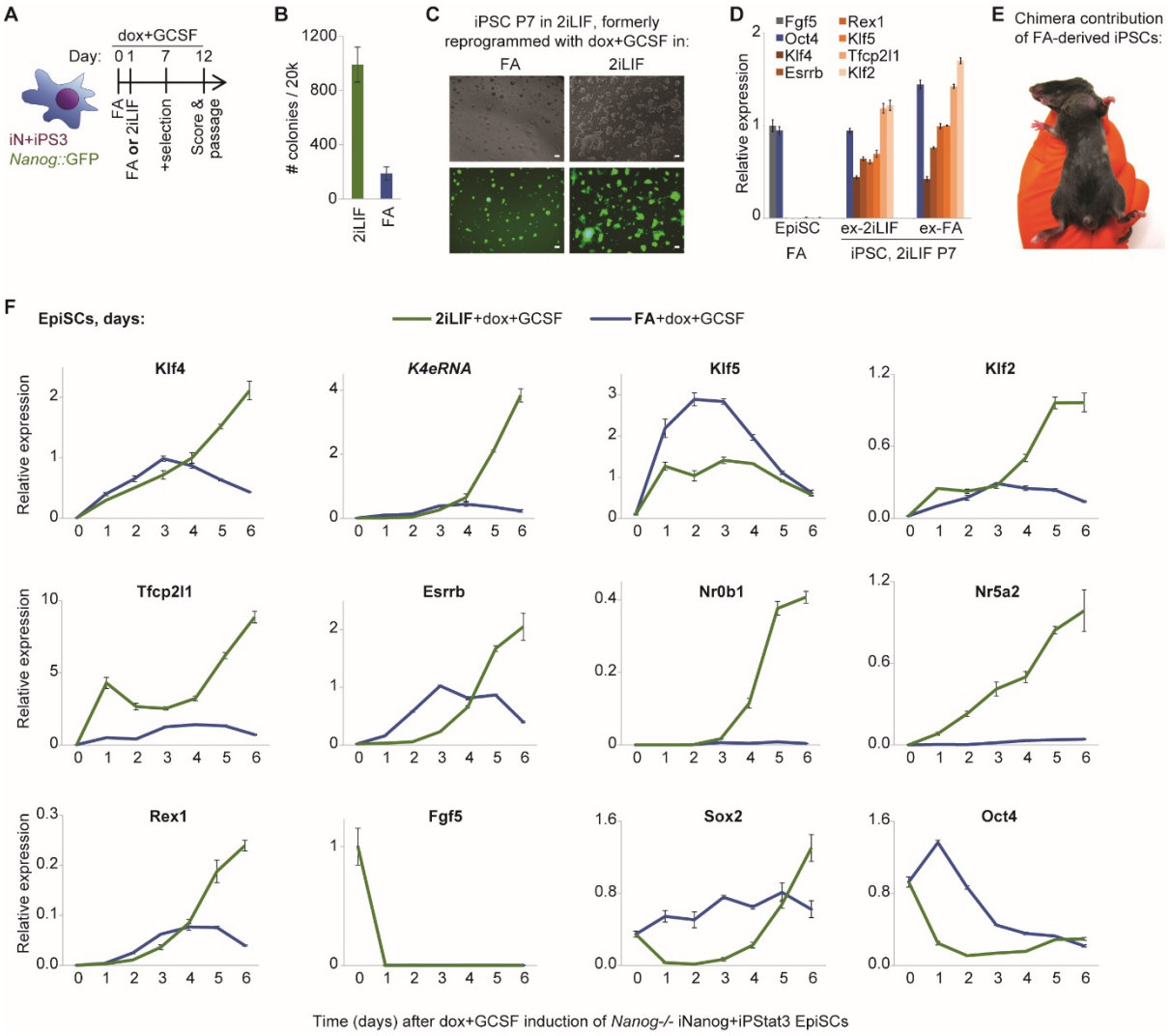


Figure 4.2.1: Reprogramming gene expression profiles in FA or 2iLIF

(A) Schematic depicting the reprogramming of *Nanog-/-* iNanog+iPStat3 (iN+iPS3) EpiSCs driven by dox+GCSF in FA or 2iLIF. *Nanog-/-* EpiSCs contain *βgeo* and *GFP* reporters under the *Nanog* promoters, and so G418 was added at day 7 to select for endogenous *Nanog* promoter activity. Since *Nanog* promoter can also be active in EpiSC FA conditions, additional selection for naïve cells was added at day 7 by applying inhibitors of Fgf and Tgfβ signalling (PD173074 and A83-01 respectively). On day 12, GFP+ colonies were scored, then passaged into 2iLIF+G418 naïve-specific conditions.
(B) Mean number of GFP+ colonies at day 12 per 20,000 cells plated ± SD (n=3).
(C) Representative images taken on passage 7 in 2iLIF+G418. Scale bars: 100 μm.
(D) RT-qPCR analysis of gene expression on passage 7 in 2iLIF+G418, compared to parental EpiSCs in FA. Mean gene expression ± SD (n=3) is shown relative to Gapdh and normalised to wild-type ESC level, except Fgf5 which is normalised to control EpiSCs.
(E) Chimeric mouse obtained after injecting *Nanog-/-* iNanog+iPStat3 iPSCs into C57/BL6 blastocysts, following reprogramming from EpiSCs in FA+dox+GCSF. Agouti coat colour indicates chimeric contribution (MF1/129 background).
(F) RT-qPCR analysis of *Nanog-/-* iNanog+iPStat3 EpiSCs following induction with dox+GCSF in FA or 2iLIF over 6 days. Mean expression was measured relative to Gapdh and normalised to FCS+LIF ESC-level =1, ± SD (n=3). Mariana Alves provided technical assistance.

I tested whether delaying the signal switch caused a reduction or increase in reprogramming, by delaying the drop in Oct4/Sox2 expression (Figure 4.2.2A–B). Interestingly, the longer the cells spent in FA+dox+GCSF prior to 2iLIF switch, the higher the reprogramming efficiency. This suggests that delaying the Oct4/Sox2 drop is beneficial, consistent with hypothesis (2), although greater proliferation in FA than 2iLIF may exaggerate the effect. By timecourse gene expression analysis, I found that the Oct4/Sox2 drops were prompted by addition of 2iLIF rather than withdrawal of FA (Figure 4.2.2C), and that switching from FA to 2iLIF on day 1, 2 or 3 elicited a similar degree of Oct4/Sox2 suppression by the following day (Figure 4.2.2D). Inverse trends were exhibited by Tfcp2l1 and Nr5a2, which are naïve-specific and positive responders to 2iLIF (Figure 4.2.1F). It would be of future interest to ascertain whether this reflects concerted yet opposite responses of Oct4/Sox2 vs Tfcp2l1/Nr5a2 to the 2iLIF signals, or whether there is direct negative regulation between them.

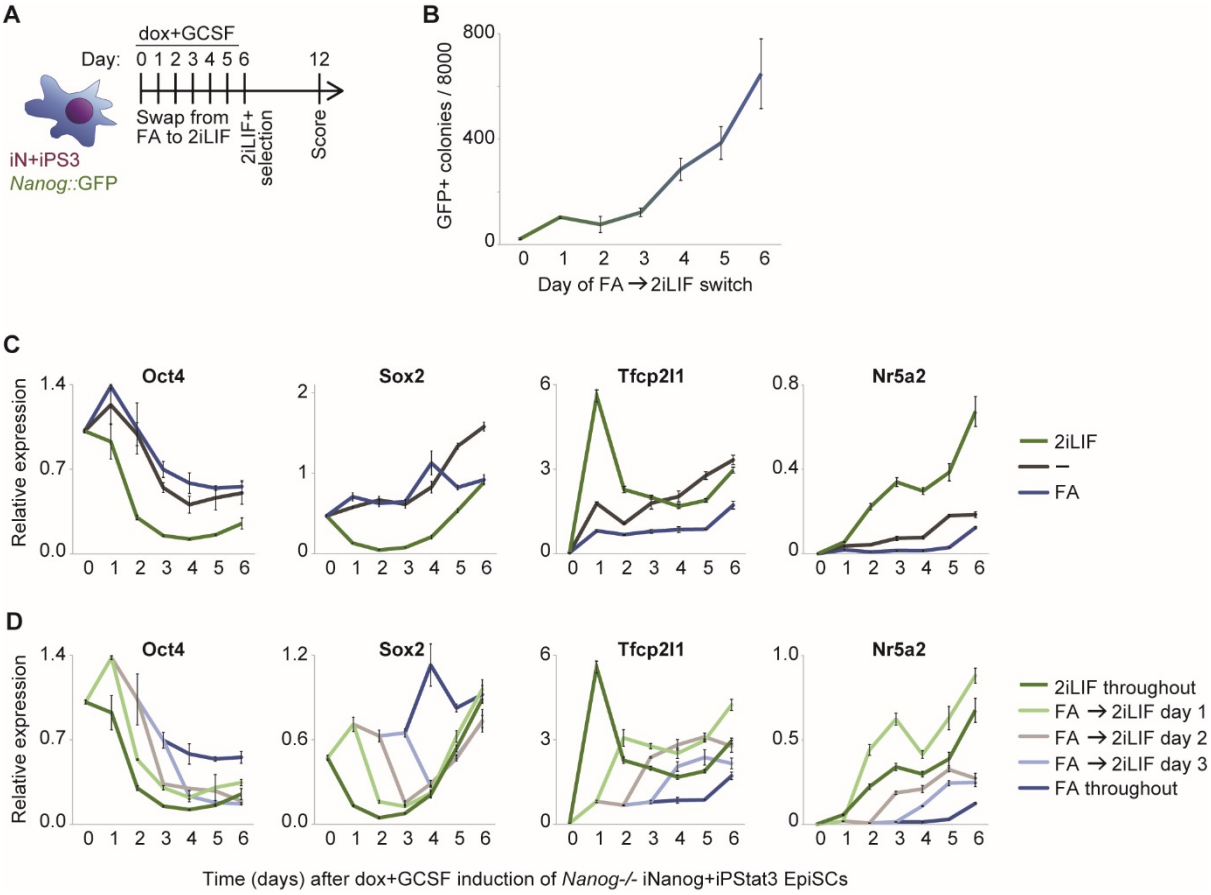


Figure 4.2.2: Network modulation by timing of signal switch in reprogramming

(A) Schematic depicting the reprogramming of *Nanog-/-* iNanog+iPStat3 (iN+iPS3) EpiSCs driven by dox+GCSF. Conditions were swapped from FA to 2iLIF on each day from 0–5, with all in 2iLIF from day 6. G418 was added at day 6 to select for endogenous *Nanog* promoter activity, together with additional selection for naïve cells using inhibitors of Fgf and Tgfβ signalling (PD173074 and A83-01 respectively). On day 12, GFP+ colonies were scored.
(B) Mean number of GFP+ colonies at day 12 per 8000 cells plated ± SD (n=3).
(C–D) RT-qPCR analysis of *Nanog-/-* iNanog+iPStat3 EpiSCs following induction with dox+GCSF in FA, basal N2B27 medium (-) or 2iLIF over 6 days **(C)**, or with medium switched from FA to 2iLIF on day 1, 2 or 3 **(D)**. Mean expression was measured relative to Gapdh and normalised to FCS+LIF ESC-level =1, ± SD (n=3).

Whilst the observations in Figures 4.2.1–2 exposed a fascinating puzzle, unfortunately the analysis of bulk averages cannot resolve the mechanistic meaning of these curious expression responses. This is further confounded by the disparate reprogramming efficiencies in FA vs 2iLIF (Figure 4.2.1B). Therefore, I devote the remainder of this Chapter to development of defined and tractable reprogramming systems of high efficiency, in which isolation of productively transitioning reprogramming intermediates is possible. Further analyses would also be better conducted in a wild-type rather than *Nanog-/-* background, so that Nanog transgene need only be used when desired rather than required for reprogramming in 2iLIF.

4.2.2 Isolation of productive reprogramming intermediates

Reprogramming is a heterogeneous and asynchronous process, in which cells undergoing the identity transition of interest are a minority subpopulation (Figure 4.1.1). As highlighted above, this presents a major challenge to mechanistic study, since bulk averages are confounded by unproductive events. To study the molecular events specifically in reprogramming cells, there are two possibilities: either develop a 100% efficient and synchronous system, or find a way to isolate the productive subpopulation. If possible at all, achieving the former would likely require a large number of inputs and manipulations, thus precluding the ability to conduct controlled single-variable experiments. Therefore, in order to understand how each factor and signal contributes to elicit identity change, the latter strategy is preferable.

I first considered which genes might provide an appropriate reporter of reprogramming progression. Reporters that have been previously used to select for iPSCs following EpiSC reprogramming include Oct4 (Guo et al., 2009) and Nanog (Festuccia et al., 2012; van Oosten et al., 2012). However, Nanog and Oct4 are expressed in both EpiSCs and iPSCs and thus are not specific to the naïve state (Section 1.2.8). Whilst Nanog or Oct4 activation in addition to the ability to clonally expand in 2iLIF are indicative of the naïve identity when considered together, such reporters do not permit accurate analysis of naïve pluripotency establishment in more permissive environmental conditions such as subsets of 2iLIF (Figure 3.2.10C). This renders Nanog and total-Oct4 reporters unsuitable for the study of signal and TF interactions during reprogramming. I attempted reprogramming using Oct4-ΔPE naïve-reporter EpiSCs (kind gift from the Surani lab) but unfortunately their reprogramming efficiencies were untenably low in my hands for unknown reasons. Furthermore, isolation of cells based on Oct4-ΔPE reporter would preclude unbiased analysis of Oct4 expression behaviour during the transition.

Considering the gene expression kinetics in Figures 3.2.9 and 4.2.1F, Rex1 appeared a suitable reporter candidate. Rex1 is expressed in naïve ESCs/iPSCs but not in EpiSCs, thus offering specificity for the destination identity. Its upregulation during initiation of EpiSC reprogramming positively correlated with future reprogramming efficiency, and occurred early enough for intermediate isolation to be feasible (unlike, for example, late-responding Nr0b1). Importantly, during ESC maintenance and differentiation, Rex1 has been extensively characterised as a sensitive yet passive proxy of naïve network strength (Kalkan et al., 2017).

To obtain *Rex1* reporter EpiSCs, *Rex1*::dGFP.IRES.bsd homozygous stud mice were crossed with wild-type females and EpiSCs derived from the resultant E6.5 embryos (Figure 4.2.3A).

These *Rex1::*dGFP reporter EpiSCs thus contain monoallelic destabilised GFP under the control of endogenous *Rex1* promoter. In order to assess whether *Rex1::*dGFP reporter activation demarcates productive reprogramming intermediates, I employed dox-inducible Klf2.2A.Nanog (K2N) as the reprogramming driver. The choice of K2N here was in part due to ongoing toxicity issues with GCSF/GY118F-forced Stat3-activation as the driver, which I resolve in the next section. Furthermore, K2N is a known efficient driver (dos Santos et al., 2014) suitable for proof-of-principle system validation, and had recently been reported as able to revert conventional primed human ESCs to a novel naïve state (Takashima et al., 2014) in a fascinating parallel to murine primed EpiSC reprogramming to naïve iPSCs.

Following *Rex1::*dGFP iK2N EpiSC reprogramming induction with 2iLIF+dox, sorting for emergent *Rex1*::dGFP reporter activation enriched for the productive subpopulation (Figure 4.2.3B–D). Importantly, this subpopulation represented intermediates that were destined to form iPSCs, but that had not already done so: sorted intermediates expressed lower levels of naïve markers than established iPSCs, and were still dox-dependent for clonogenicity in 2iLIF (Figure 4.2.3C–F). Naïve gene reactivation was higher in the dGFP+ intermediates, and progressed over time in conjunction with gradual Rex1 upregulation (Figure 4.2.3F). Together, this indicates that *Rex1::*dGFP reporter can be used to resolve productive intermediates from the average, rendering transitioning events available for mechanistic study.

Again, both Oct4 and Sox2 dropped in the bulk average following reprogramming induction in 2iLIF (Figure 4.2.3F). Strikingly, Oct4 and Sox2 behaviour resolved in opposite manners in the *Rex1::*dGFP+ subpopulation: Oct4 was maintained at or slightly above EpiSC/ESC expression level throughout the productive transition, whereas Sox2 was suppressed more in the dGFP+ fraction. This highlights the importance of studying productive rather than bulk populations in order to understand the logic of cell identity transitions, and it will be of great interest to ascertain whether these opposing Oct4/Sox2 responses are K2N-specific or a general feature of productive reprogramming.

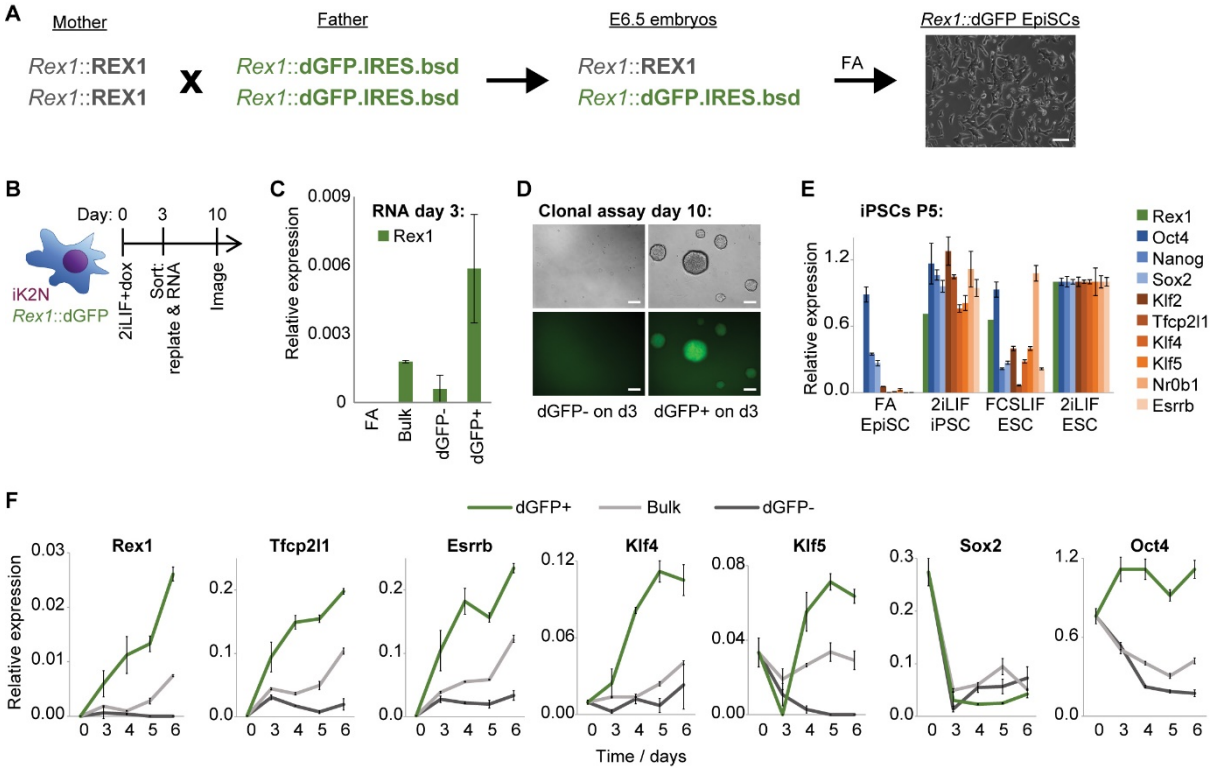


Figure 4.2.3: Validation of *Rex1::*dGFP reporter for isolation of productive intermediates

(A) Schematic of the strategy used to generate *Rex1+/dGFP.IRES.bsd* (*Rex1::*dGFP) EpiSCs. Wild-type females were crossed with homozygous males, and heterozygous EpiSCs derived from resultant E6.5 embryos. All alleles are under control of the endogenous *Rex1* promoter. dGFP has a half-life of 2 hours. Bsd confers resistance to blasticidin. Scale bar: 100 μm.
(B) Reprogramming of *Rex1::*dGFP iKlf2.2A.Nanog (iK2N) EpiSCs was induced by 2iLIF+dox. Emergent dGFP+ cells were isolated by FACS on day 3 and analysed by RT-qPCR and clonal assay in 2iLIF+dox.
(C) RT-qPCR analysis of Rex1 expression in emergent dGFP+, dGFP- and bulk sorted populations at day 3. This Rex1 expression is from the wild-type allele of the heterozygous cells, and positively correlates with dGFP driven by the other *Rex1* promoter. Mean expression is shown relative to Gapdh then normalised to wild-type ESC level=1, ± SD (n=3).
(D) Representative phase and dGFP images were taken on day 10, after clonal assay on cells sorted at day 3. Scale bars: 100 μm.
(E) RT-qPCR analysis of gene expression of *Rex1::*dGFP iK2N iPSCs at passage 5, together with parental EpiSCs and wild-type ESCs. Mean expression is shown relative to Gapdh then normalised to wild-type 2iLIF ESC level=1, ± SD (n=3).
(F) RT-qPCR analysis of gene expression in emergent dGFP+, dGFP- and bulk sorted populations on days 3–6. Mean expression is shown relative to Gapdh then normalised to wild-type 2iLIF ESC level=1, ± SD (n=3).

4.2.3 Optimisation of Stat3 activation by GCSF/GY118F

PStat3 is one of the most important drivers to include in the study of EpiSC reprogramming. PStat3 is one of very few naïve factors that is essential for pluripotency establishment *in vivo* (Do et al., 2013) (Figure 3.1.1). Furthermore, it is to date the only known single factor to dominantly instruct naïve pluripotency whilst maintaining primed-specifying FA culture conditions (van Oosten et al., 2012). Unlike Oct4, PStat3 is rapidly lost in the peri-implantation epiblast, indicative of a naïve-specific role in the pluripotency continuum. Elegantly, provision of PStat3 as the driving TF overlaps with one of the three exogenous signals instructing naïve pluripotency: Stat3 is activated by the LIF component of 2iLIF. Therefore, whilst other TF drivers provide an additional input on top of 2iLIF, forced Stat3 activation together with PD03 and Chiron yields only three variables to dissect in terms of input to cellular identity computation. However, as alluded to above, I repeatedly noticed some toxicity issues associated with GCSF treatment of GY118F-containing EpiSCs of various genetic backgrounds. This was problematic both in terms of efficiency reduction, and contamination of early timepoint samples with dying material. Therefore, methodology for forcing Stat3 activation merited optimisation.

In order to activate Stat3 in EpiSCs, provision of LIF is insufficient since EpiSCs have a limited ability to transduce LIF signal via LIFR/GP130 (Yang et al., 2010). Treatment with IL6 and soluble IL6 receptor could be an alternative, but would nevertheless rely on endogenous GP130 and would stimulate Mek/Erk signalling in addition to Jak/Stat3. Since GCSF-stimulation of the GY118F receptor transgene elicits specific and sustained Stat3 activation (Burdon et al., 1999) (Figure 1.2.4), it remained the most promising approach. Anecdotally, in preliminary reprogramming experiments driven by iNanog+GY118F.dsRed from a heterogeneous starting EpiSC population, I noticed that iPSCs generated with or without GCSF exhibited lower or higher levels of dsRed respectively (Figure 4.2.4A). This suggested to me that the level of GY118F.dsRed might be important, with too much being incompatible with reprogramming and/or survival following GCSF-stimulation. In contrast, perhaps a degree of leak could explain the preference for high GY118F.dsRed when reprogramming was driven by dox alone. There is precedent in the literature for the level of PStat3 playing an important role in cell fate decisions. For example, Stat3 activation above a certain threshold can induce ESCs to differentiate towards the trophectoderm lineage, rather than promoting self-renewal (Tai et al., 2014). Therefore, I asked whether a particular level of PStat3 was optimal for naïve pluripotency specification during reprogramming, and if so whether I could achieve this by defined modulation of the GCSF/GY118F system.

I first took advantage of established *Stat3+/+* and *Stat3-/-* ESCs to investigate how levels of GY118F, Stat3 and PStat3 pertain to naïve pluripotency maintenance. In the ESC context, cellular identity is stably established and thus controlled manipulation of factors can provide information regarding compatibility with the naïve identity. Rather than generating an array of clonal lines, I instead let the ESCs inform me as to their preferred PStat3 level, then applied this lesson to reprogramming system design. I transfected *Stat3+/+* and *Stat3-/-* ESCs with transgenes driving constitutive Stat3 and/or GY118F expression in a variety of combinations and ratios, then selected stable transgenic lines both in the presence and absence of LIF/GCSF stimulation (Figure 4.2.4B). Subsequently, I performed acute induction of previously unstimulated populations, and compared resultant PStat3 levels with those exhibited following 3 weeks of self-selection (Figure 4.2.4C–E). Whilst acute inductions elicited high PStat3 levels, long-term self-selection generally yielded PStat3 levels that were the same as those exhibited in wild-type 2iLIF ESC cultures (Figure 4.2.4 red stars). Therefore, the preferred PStat3 dosage for naïve pluripotency maintenance is the endogenous level. This is in sharp contrast to other naïve factors known to drive reprogramming of EpiSCs into iPSCs, such as Nanog, Esrrb, Tfcp2l1 and Klfs, whose overexpression is beneficial for ESC self-renewal (Martello and Smith, 2014). Even when total Stat3 protein is grossly overexpressed (Figure 4.2.4 purple circles), ESCs have the capacity to precisely restrict the phosphorylated subset.

Further insights were obtained by careful observation of ESC cultures during this self-selection process (Figure 4.2.4F–G). GCSF stimulation of GY118F.dsRed was highly detrimental in a *Stat3-/-* background, causing a great deal of death from which a surviving population eventually emerged with no visible dsRed. Meanwhile, GCSF-stimulated *Stat3+/+* GY118F.dsRed ESCs self-selected a low level of GY118F following a period of cell death (Figure 4.2.4F), in the same manner as iPSCs reprogrammed by GCSF/GY118F (Figure 4.2.4A). Interestingly, *Stat3-/-* ESCs constitutively expressing Stat3 but not GY118F did not die in response to LIF stimulation, but initially exhibited a high frequency of spontaneous differentiation (Figure 4.2.4G). Together, this suggests that the naïve pluripotent identity is compatible with levels of PStat3 below a particular threshold, consistent with the work of Tai et al., 2014. Above this, cells appear to die if PStat3 activation is driven by Socs3-insensitive GCSF/GY118F, or differentiate if PStat3 is driven by Socs3-sensitive LIF/LIFR.

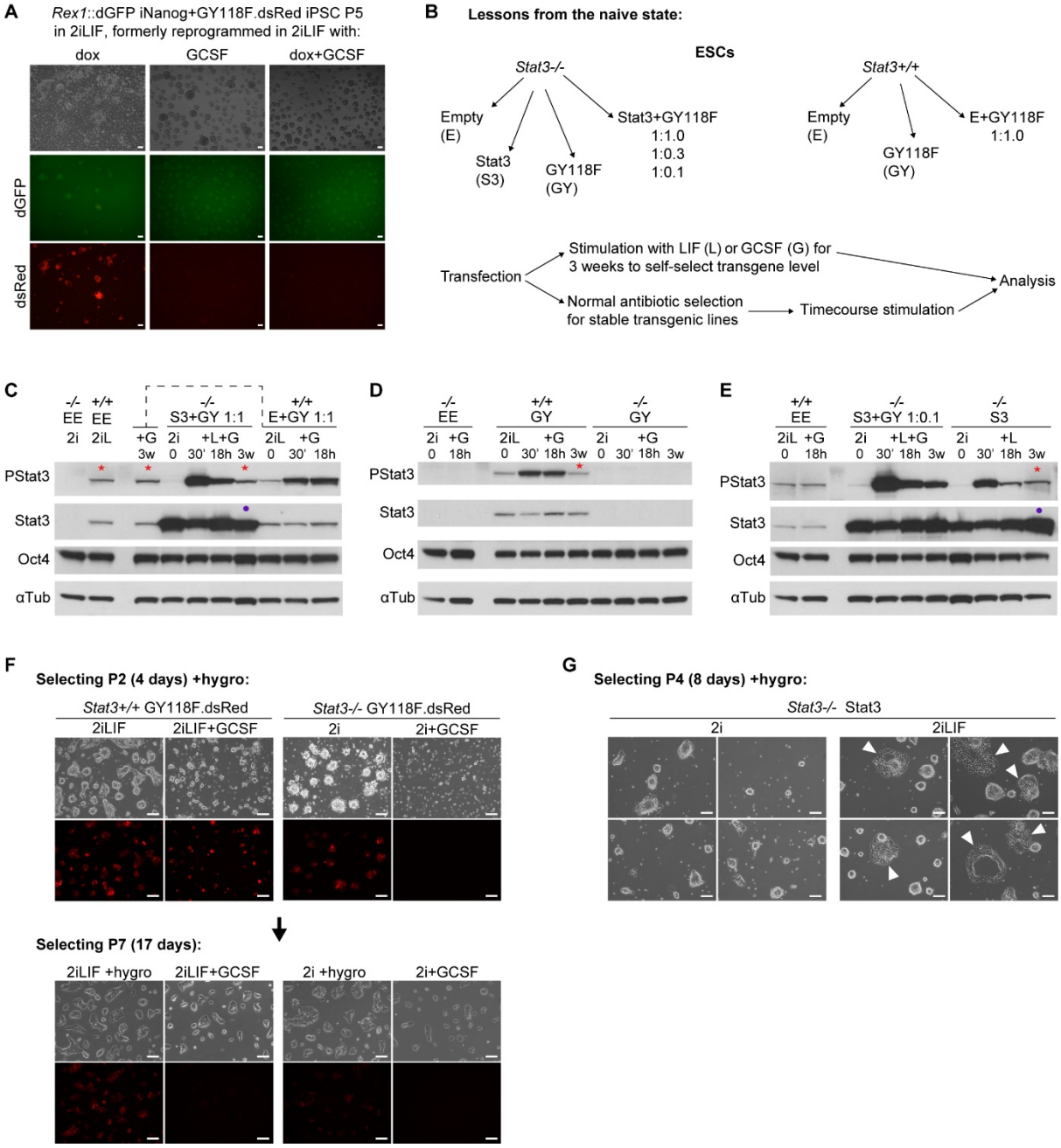


Figure 4.2.4: Lessons from the naïve state regarding appropriate Stat3 activation level

(A) Representative images of passage 5 iPSCs in 2iLIF, generated by reprogramming of *Rex1*::dGFP iNanog+GY118F.dsRed EpiSCs in 2iLIF±dox±GCSF. Scale bars: 100 µm.
(B) Experimental strategy to ascertain the preferred PStat3 level in ESCs. Numbers refer to the µg of plasmid transfected per 250,000 cells in a 6well.
(C–E) Western blot analysis[2] of PStat3, total Stat3, Oct4 and αTubulin following stimulation with LIF or GCSF for 30 minutes, 18 hours or 3 weeks (3w). Key: see panel B. The same membrane was stripped between PStat3 and total Stat3 blots. Apologies for the error in loading order in panel C (dashed line).
(F–G) Representative phase and dsRed images during the selection process in ESCs. Scale bars: 100 µm.

[2] Matched RNA samples are in the Silva lab -80 for anyone who would like to continue this work. Additionally, EpiSCs differentiated from the 3w self-selected ESCs are in the Silva lab cell bank.

Since stimulation of PStat3 above endogenous level appears incompatible with naïve pluripotency maintenance, it could plausibly also be suboptimal for naïve identity establishment during reprogramming. Perhaps this accounts for the toxicity I was observing during GCSF/GY118F-mediated reprogramming, and the prior reporting of GY118F-driven reprogramming efficiencies at 1–2% (Yang et al., 2010). Assessment of the relationships between GY118F, Stat3 and reprogramming first necessitated the generation of *Stat3-/-* and matched control EpiSCs. Therefore, I differentiated *Stat3-/-* and littermate *Stat3+/+* ESCs into EpiSCs for 8 passages in FA, then transfected them with various combinations and ratios of Stat3 and GY118F transgenes (Figure 4.2.5A). After 5 passages of selection for stable transgene integration without induction, I tested their reprogramming propensity in 2i with LIF/GCSF. Since *Stat3-/-* and littermate *Stat3+/+* cells do not contain any pluripotency reporter, it is not possible to select for or quantify reprogramming in the usual manner. Instead, after 4 days of reprogramming I passaged all cells 4 times in 2iLIF with thorough dissociation to single cells, applying additional inhibitors of Fgf (PD173074) and Tgfβ (A83-01) signalling to eliminate unreprogrammed material that should not survive such conditions. The number of colonies at passage 4 thus gives a semi-quantitative reflection of reprogramming efficiency, representative images of which are shown in Figure 4.2.5B–C.

Despite their origin as differentiated ESCs rather than embryo-derived EpiSCs, *Stat3+/+* and *Stat3-/-* Stat3-rescued EpiSCs were not permissive for reprogramming in only 2iLIF (Figure 4.2.5B), indicating that a full conversion to 'late-stage' EpiSCs had been achieved by the differentiation (Han et al., 2010). GCSF/GY118F rescued reprogramming of *Stat3+/+* but not *Stat3-/-* EpiSCs, formally demonstrating that GY118F drives reprogramming in a Stat3-dependent manner. Reprogramming did occur for *Stat3-/-* Stat3+GY118F EpiSCs. Intriguingly, GCSF stimulation generated iPSCs at higher efficiency when a lower dose of GY118F had been transfected (Figure 4.2.5C 2i+GCSF). Conversely, leaky reprogramming without GCSF occurred at a higher efficiency with a higher dose of GY118F (Figure 4.2.5C 2iLIF). Together, these observations are consistent with the hypothesis that too much PStat3 is detrimental to reprogramming, yet also demonstrated that some PStat3 was required. The latter was confirmed by RT-qPCR analysis of naïve gene expression after 4 days of reprogramming: prevention of Oct4 loss together with upregulation of naïve genes Klf4 and Rex1 occurred only in the presence of activated Stat3 (Figure 4.2.5D).

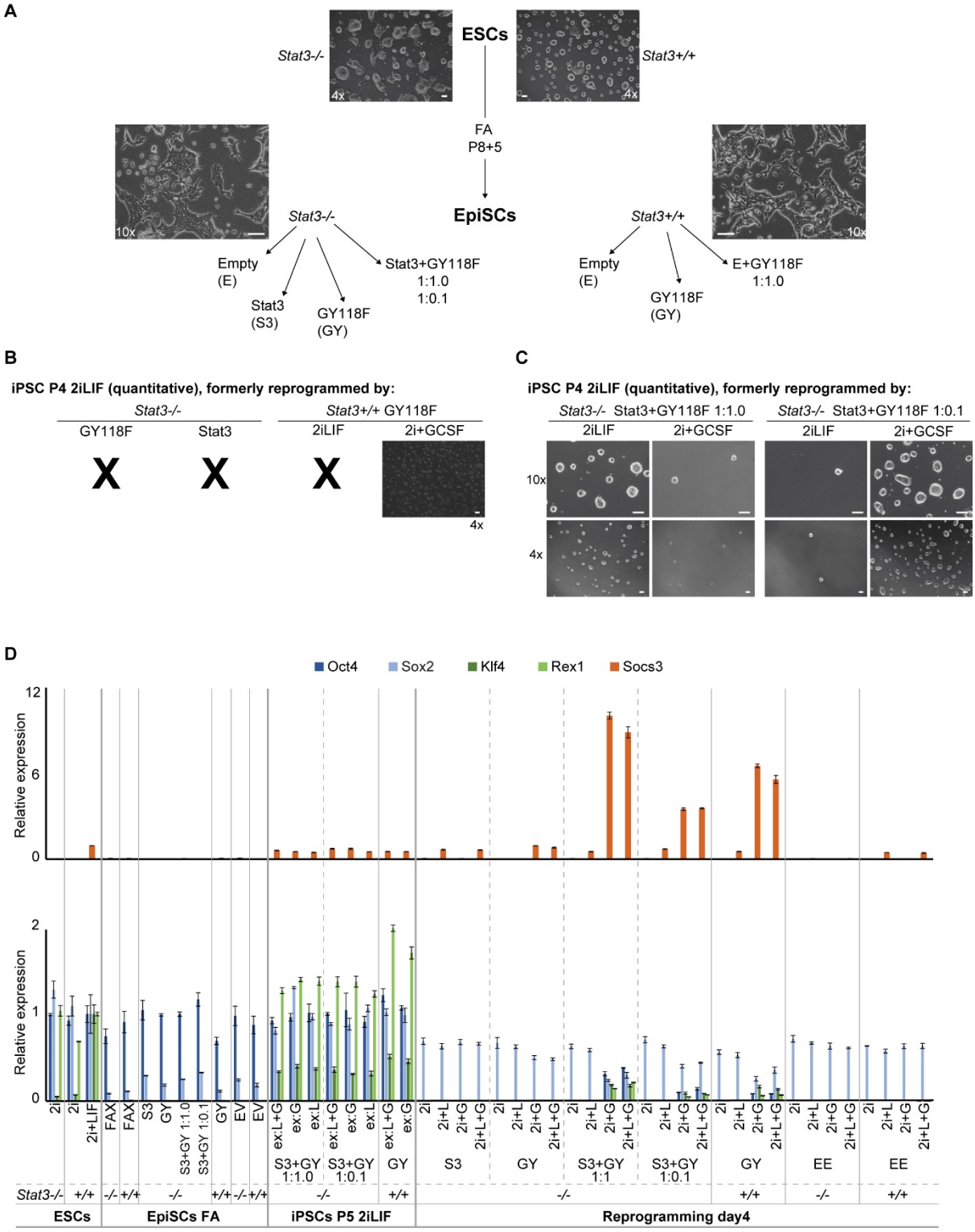


Figure 4.2.5: Balance between Stat3 and GY118F in EpiSC reprogramming

(A) *Stat3-/-* and littermate *Stat3+/+* ESCs were differentiated to EpiSCs by 8 passages in FA. Representative phase images of ESCs and resultant EpiSCs are shown. The EpiSCs were then transfected as indicated, and stably selected for 5 passages. Numbers refer to the µg of plasmid transfected per 250,000 cells in a 6well. Scale bars: 100 µm.

(B–C) Representative phase images of iPSCs after 4 passages of all material in 2iLIF with additional inhibitors of Fgf (PD173074) and Tgfβ (A83-01) signalling to eliminate unreprogrammed material. The number of colonies thus reflects reprogramming efficiency. Scale bars: 100 μm.
(D) RT-qPCR analysis of gene expression in parental ESCs, EpiSCs after differentiation then transfection, day 4 of reprogramming from EpiSCs, and resultant iPSCs. Mean expression is shown relative to Gapdh then normalised to *Stat3+/+* 2iLIF ESC level=1. Key: see panel A.

To conclusively demonstrate a causal relationship between PStat3 induction level and reprogramming propensity, clonal lines with different transgene amounts should be generated and compared side-by-side. Despite many attempts, I was never able to generate clonal transgenic lines from EpiSCs by either picking or sorting. On the rare occasion that a clone expanded, its reprogramming ability was permanently compromised. This is not surprising, since EpiSCs are epithelial and known not to tolerate clonal passaging, yet it unfortunately limited further advancement of this line of inquiry. Nevertheless, the circumstantial evidence described in this Section 4.2.3 gave sufficient grounds to test transfection of *Rex1::*dGFP EpiSCs with a low dose of GY118F as an improved tool for PStat3-driven reprogramming.

I transfected embryo-derived *Rex1::*dGFP EpiSCs (Section 4.2.2) with high and low amounts of GY118F transgene: 1.0 or 0.1 μg per 250,000 cells, subsequently selected with 150 or 50 μg/ml hygromycin respectively. Both hygromycin concentrations killed untransfected controls. The reprogramming efficiency and kinetics of the low GY118F line were much greater than the high GY118F (Figure 4.2.6A). This reprogramming was, as expected, entirely GCSF-dependent (Figure 4.2.6A–B). Importantly, the level of PStat3 induced by each line positively correlated with the amount of GY118F transfected, and negatively correlated with its reprogramming propensity (Figure 4.2.6C–D). I had considered further manipulation of *Rex1::*dGFP EpiSCs to include Stat3 transgene as well as GY118F, in case the EpiSC level of Stat3 was too low to respond optimally to GCSF/GY118F and in light of promising results in the *Stat3-/-* Stat3+GY118F 1:0.1 EpiSCs of Figure 4.2.5. However, endogenous Stat3 was quickly upregulated in response to GCSF/GY118F (Figure 4.2.6D), indicating that EpiSCs were competent to implement the positive feedback loop of PStat3 on Stat3 expression previously observed in ESCs (Figure 3.2.1) (Bourillot et al., 2009; Martello et al., 2013; van Oosten et al., 2012). Since low GY118F was sufficient for excellent reprogramming performance, and to avoid unnecessary perturbations, I continued without additional Stat3 transgene in subsequent work.

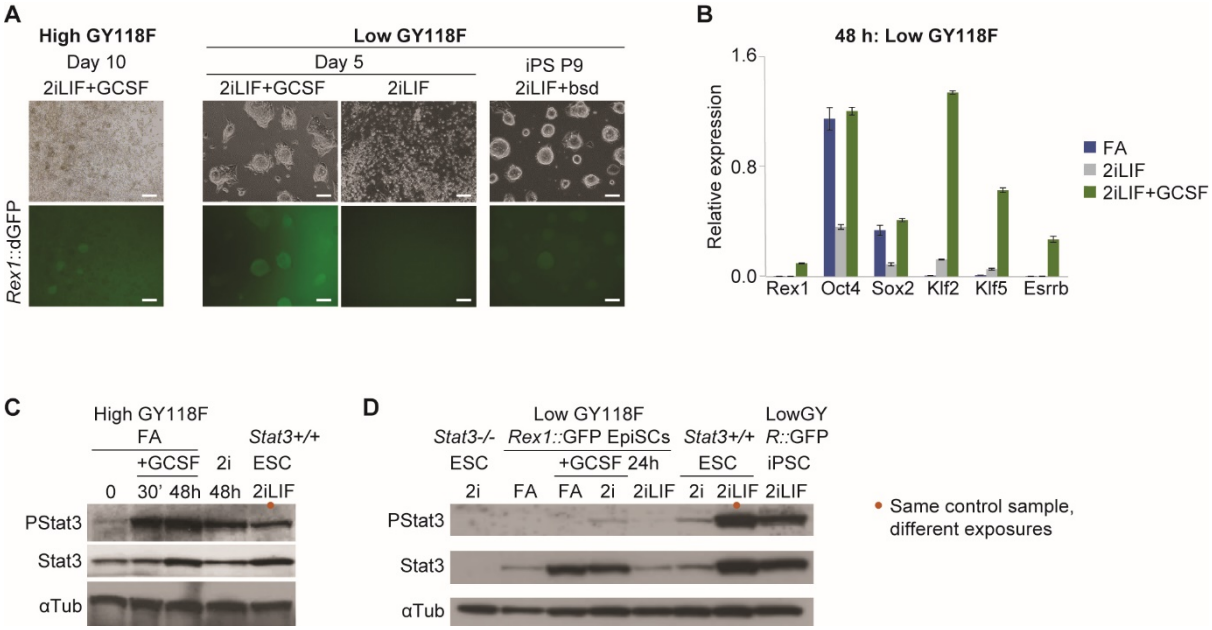


Figure 4.2.6: High-efficiency low-GY118F reprogramming

(A) Representative phase dGFP images of *Rex1::*dGFP high GY118F EpiSCs on day 10 of reprogramming, low GY118F on day 5 and resultant iPSCs. High refers to transfection with 1.0 µg of GY118F plasmid per 250,000 cells, while low refers to 0.1 µg. Scale bars: 100 µm.
(B) RT-qPCR analysis of gene expression 48 hours after reprogramming induction of low GY118F EpiSCs with 2iLIF±GCSF, or maintenance in FA. Mean expression is shown relative to Gapdh then normalised to wild-type 2iLIF ESC level=1, ± SD (n=3).
(C–D) Western blot analysis of PStat3, total Stat3 and αTubulin in high and low GY118F *Rex1::*dGFP EpiSCs, following stimulation with GCSF in the indicated conditions. The same membrane was stripped between PStat3 and total Stat3 blots. These western blots were performed by Petros Fessas.

4.2.4 Further technical improvements to EpiSC culture and reprogramming

Prior to returning to the study of reprogramming mechanisms, I made a few further improvements to technical aspects of EpiSC culture and reprogramming. Since the metabolism of EpiSCs is largely glycolytic rather than oxidative (Zhou et al., 2012), I tested whether EpiSC maintenance and reprogramming could be improved under hypoxic conditions. Indeed, culture at 5% rather than atmospheric oxygen levels improved EpiSC morphology (Figure 4.2.7A), reprogramming efficiency and kinetics (Figure 4.2.7B). From hereon, all EpiSC cultures and reprogramming experiments are performed in low oxygen conditions.

At the time of this work, new publications indicated that addition of XAV939 (XAV) improved EpiSC cultures by reducing heterogeneity, spontaneous differentiation and lineage marker expression (Kim et al., 2013; Sugimoto et al., 2015; Sumi et al., 2013). XAV is a small molecule that attenuates Wnt signalling, although its mechanisms of action in EpiSCs remain incompletely understood. Like Fgf signalling in naïve ESCs, endogenous Wnt signalling in EpiSCs promotes spontaneous differentiation, lineage priming, and heterogeneity (Kurek et al., 2015; Tsakiridis et al., 2014). This corresponds to the driving force of Wnt *in vivo* to push the epiblast towards primitive streak/mesodermal fates (Huelsken et al., 2000; Liu et al., 1999). Therefore, by inhibiting Wnt signalling, EpiSCs are held in a more homogenous state proposed to represent the 'ground state' of primed pluripotency (Morgani et al., 2017; Sumi et al., 2013). It is a further testament to the identity distinction between naïve and primed pluripotent states, that Wnt signalling promotes self-renewal while Fgf signalling triggers differentiation of the former, while the converse applies to the latter.

Work in the Nichols group confirmed that addition of XAV to FA EpiSC culture conditions improved derivation and maintenance of more homogenous EpiSC lines. Therefore, in collaboration we derived fresh *Rex1::*dGFP EpiSCs under hypoxia in FA+XAV, which I use for the remainder of this thesis.

Together, these improvements to EpiSC culture and reprogramming should facilitate the mechanistic study of cell identity transitions in this model system. Not only is it now possible to isolate productively transitioning intermediates on the basis of *Rex1::*dGFP reporter activation (Section 4.2.2), but higher efficiencies lend more meaning to analysis of earlier events by bulk averages, and improve the technical feasibility of complex multivariate experiments.

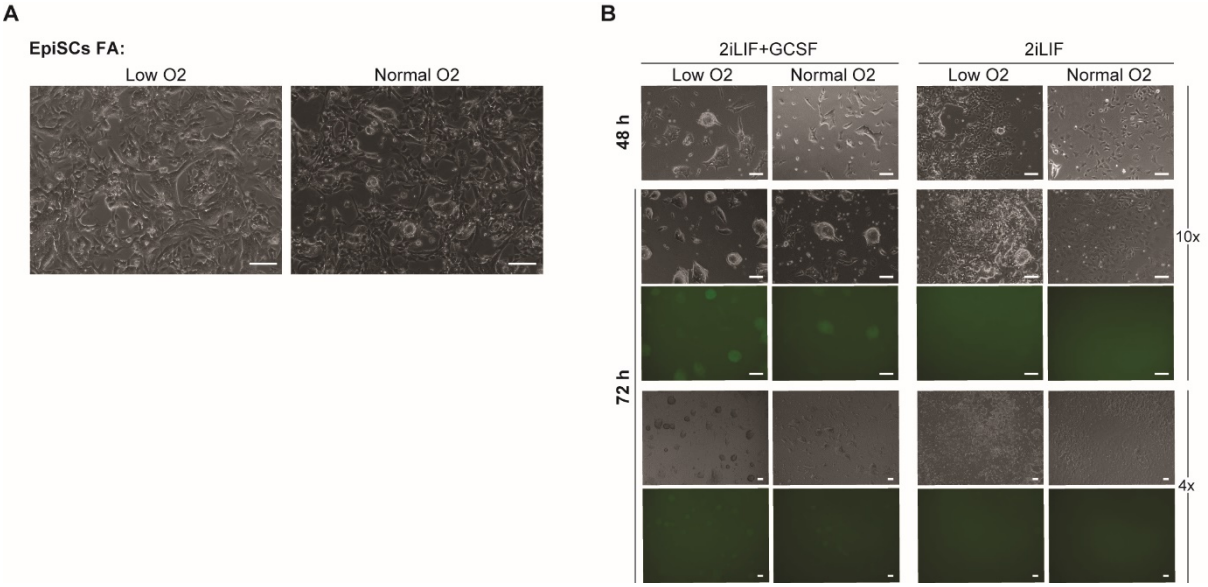


Figure 4.2.7: Improved EpiSC reprogramming in hypoxia

(A) Representative 10x phase images of EpiSCs cultured in FA in hypoxic vs normoxic conditions. This was still without XAV. Scale bars: 100 µm.
(B) Representative 10x and 4x phase and dGFP images of *Rex1*::dGFP iPStat3 EpiSC reprogramming at 48 and 72 h in hypoxic vs normoxic conditions. Reprogramming is driven by GCSF stimulation in 2iLIF. Scale bars: 100 µm.

4.2.5 Model set of single reprogramming drivers

In order to causally ascribe independent genetic and signal variables to reprogramming events, use of single drivers is necessary. Therefore, I tested reprogramming efficacy of individual naïve factors in the new FA+XAV embryo-derived *Rex1::*dGFP EpiSCs under hypoxia (Figure 4.2.8A–C). Dox-inducible (i) transgenes were used for Esrrb, Klf2, Klf4, Nanog, Tfcp2l1 and Klf5. Stat3 activation by phosphorylation (iPStat3) was elicited by GCSF-stimulation of the GY118F receptor transgene, with the optimised low GY118F dosage as established in Section 4.2.3. Use of stably-integrated inducible transgenes has several benefits, including experiment reproducibility, and induction upon demand permits analysis of direct downstream effects in a controlled manner from an uncompromised starting cell identity. This is in contrast to the majority of published EpiSC reprogramming experiments conducted with constitutively expressed transgenes or transient transfections.

Control EmptyVector+CAG.rtTA3 (EVrtTA3) did not yield any iPSCs (Figure 4.2.8B–C). This confirms that the starting cells represented 'late-stage' EpiSCs (Han et al., 2010) and that there is no background to TF-driven reprogramming. Impressively, following only 4 days of single-driver induction, reprogramming efficiencies of up to 80% were observed. This is a testament to the system optimisation of Chapter 4. Previously reported EpiSC reprogramming efficiencies with single drivers have ranged from <1–3%, while efficiencies ≤10% have been achieved with multiple drivers (Gillich et al., 2012; Yang et al., 2010). Here, substantially higher efficiencies and faster kinetics were achieved, without compromising on the number or control of variables.

iEsrrb, iPStat3 and iKlf2 were the three most efficient single drivers in 2iLIF (Figure 4.2.8C). iPSCs established from these lines had identity-appropriate gene expression signatures (Figure 4.2.8D). When analysed by unsupervised hierarchical clustering for all expressed genes, they did not cluster according to line, indicating transcriptional equivalency to each other and to ESCs (Figure 4.2.8E). Interestingly, Esrrb, PStat3 and Klf2 each inputs to the naïve network along a different regulatory axis (Figure 4.2.8F), that of Chiron, LIF and PD03 respectively (Martello et al., 2012; Niwa et al., 1998; Yeo et al., 2014). Therefore, I take iEsrrb, iPStat3 and iKlf2 as a complementary model set of single reprogramming drivers for mechanistic study.

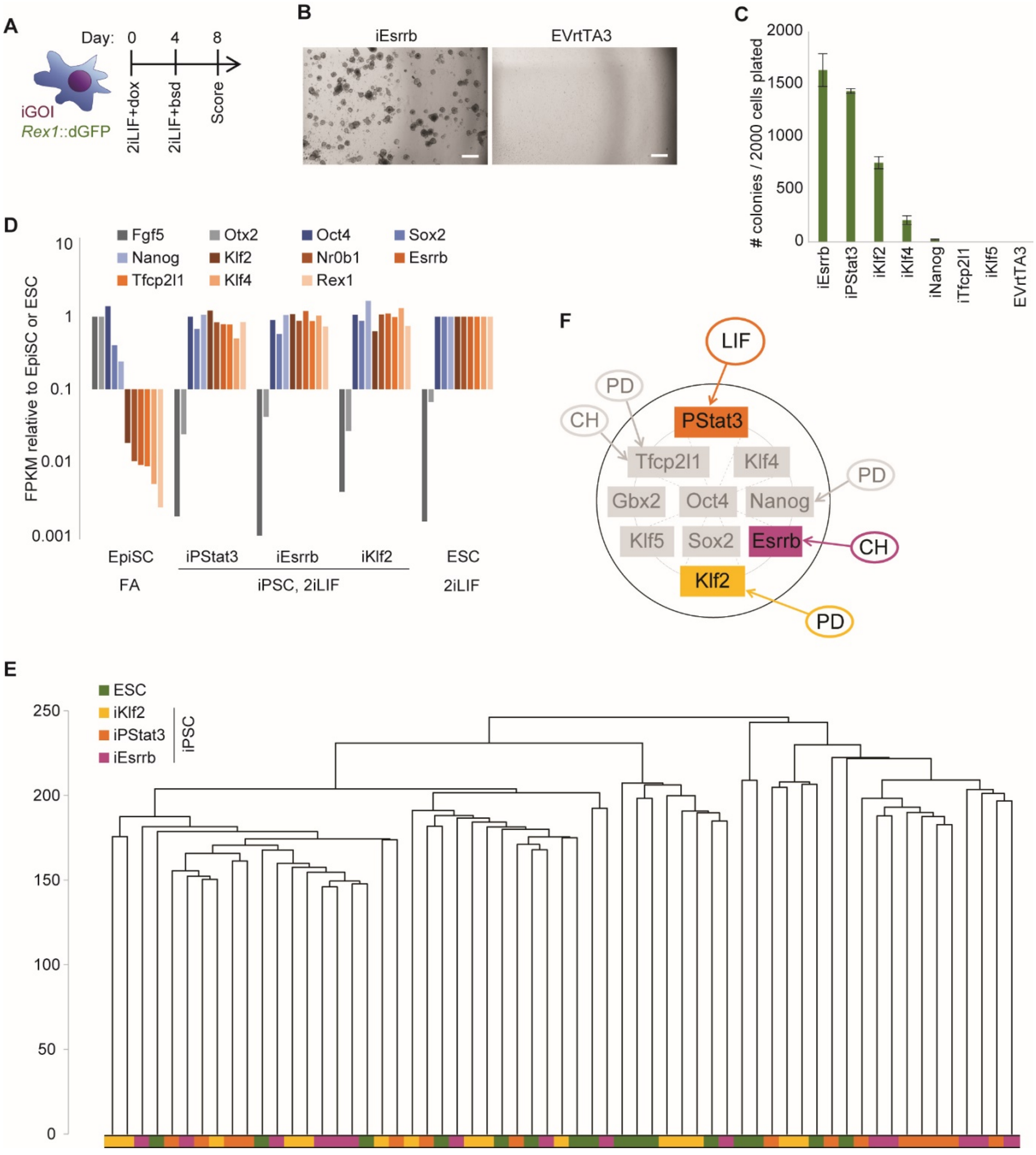


Figure 4.2.8: Model set of single reprogramming drivers

(A) Reprogramming protocol for *Rex1::*dGFP EpiSCs. iGOI: inducible driver gene-of-interest.
(B) Representative phase images of iEsrrb and EVrtTA3 wells on day 8. Scale bars: 500 µm.
(C) Mean number of iPSC colonies on day 8 ±SD (n=3) per 2000 cells plated. Note that only 4 days of transgene induction is a stringent test of driver efficacy.
(D) Mean expression of iPSCs at passage 5, together with control EpiSCs and ESCs (n=16). Greys: primed markers. Blues: core pluripotency markers. Oranges: naïve markers.
(E) Unsupervised hierarchical cluster of sequenced single cells, computed with the Ward.D2 agglomeration method and Euclidean distances, for all expressed genes (FPKM>0).
(F) Schematic depicting the inputs of Esrrb, PStat3 and Klf2 to the naïve network.
The RNAseq analyses were conducted in collaboration with Giuliano Stirparo, and colony forming assay with Tim Lohoff.

To confirm that *Rex1::*dGFP reporter activation demarcated productive intermediates for iEsrrb, iPStat3 and iKlf2 as it did for iK2N (Section 4.2.2), I isolated emergent dGFP+ reprogramming subpopulations from each driver on days 2 and 3. When replated in 2iLIF+dox/GCSF, emergent dGFP+ cells were indeed destined to form naïve colonies with efficiency comparable to ESCs (Figure 4.2.9 A–B). iPSCs thus obtained were chimera and germline competent (Figure 4.2.9C–E), demonstrating true establishment of naïve pluripotency.

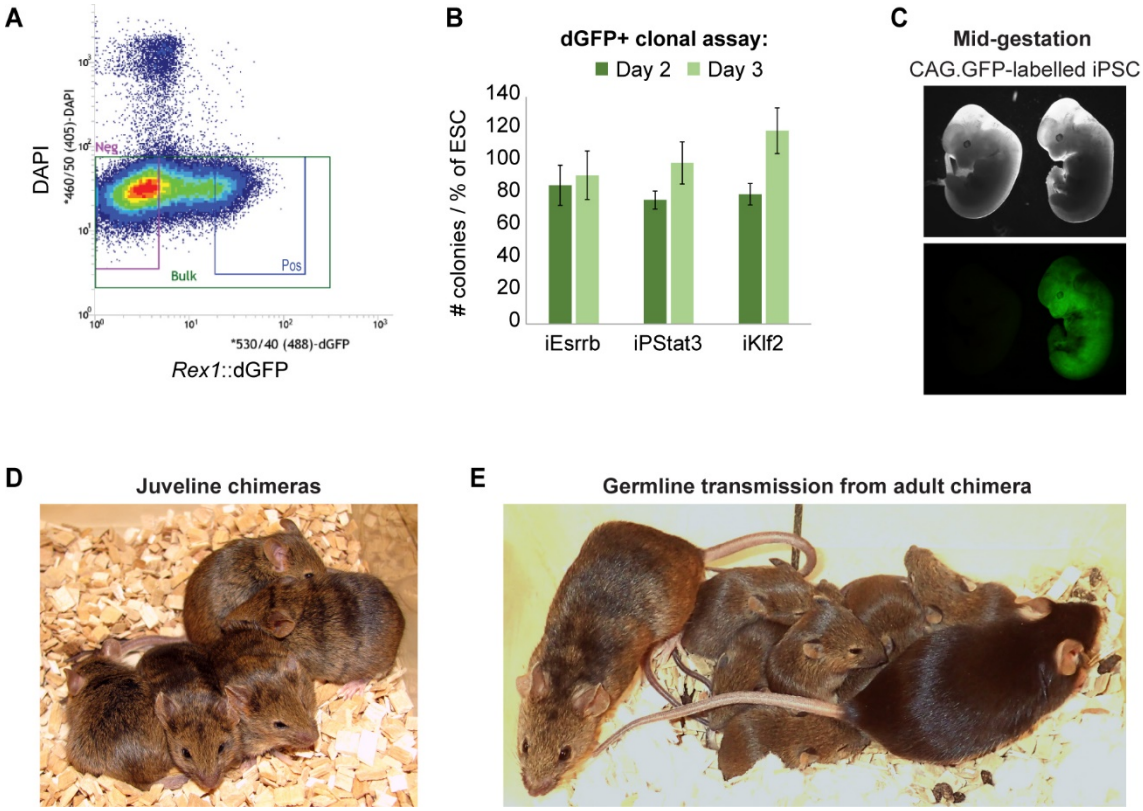


Figure 4.2.9: Functional validation of single-driver *Rex1::*dGFP EpiSC reprogramming

(A) FACS-plot showing emergence of *Rex1*::dGFP expression during reprogramming. The example here is iEsrrb 2iLIF+dox day 3. *Rex1::*dGFP ESCs and EpiSCs were used to determine positive (pos) and negative (neg) gates respectively.
(B) *Rex1::*dGFP+ cells were isolated by FACS on days 2 and 3 and plated for clonal assay: reprogramming intermediates were plated in 2iLIF+dox, established *Rex1::*dGFP ESCs in 2iLIF. Dox was withdrawn and blasticidin applied on day 6. Colonies were scored manually on day 9. Mean number of iPSC colonies ±SD (n=3) is indicated as a percentage of ESC colonies for each experiment.
(C–E) Blastocyst injection of *Rex1::*dGFP iPSCs (agouti) yielded high-contribution, germline competent chimeras.
(C) Constitutive GFP transgene was added to visualise mid-gestation chimerism.
(D) Litter of juvenile chimeras.
(E) Chimeric adult (left, father) capable of germline transmission (centre, agouti litter).

4.3 Discussion

The combined work of Chapter 4 led to the generation of a tractable identity transition model set. This comprises *Rex1::*dGFP embryo-derived EpiSCs cultured in FA+XAV under hypoxia, stably containing inducible transgenes for single reprogramming drivers. iEsrrb and iKlf2 are dox-inducible, whereas iPStat3 is achieved by applying GCSF to cells transfected with a low amount of GY118F (see Figure 1.2.4).

This meets the requirements for an identity transition model, as specified in Section 1.3.2:

a. Source and destination identities are distinct and well defined: 'ground-state' primed EpiSCs and naïve iPSCs respectively (Sections 1.2, 1.3.3, 1.3.4, 4.2.4).

b. Source and intermediate cells are amenable to perturbation.

c. Genetic and signal variables can be controlled and independently manipulated:

    i. Single driver TFs can be induced on demand using dox or GCSF (Section 4.2.5). Further genetic manipulation can be achieved using siRNA to knock-down specific genes, thanks to the amenability of EpiSCs to transfection.

    ii. Signal variables are defined and minimal, comprising only exchange of FA+XAV to 2iLIF thanks to the shared N2B27 basal medium and fibronectin coating.

    iii. Signal variables can be manipulated independently of genetic drivers whilst still achieving the identity transition of interest (Section 4.2.1) (van Oosten et al., 2012).

d. Transcriptional responses are sufficiently rapid to infer regulatory relationships and study network installation kinetics (Sections 3.2.3, 4.2.1).

e. In 2iLIF, the chosen drivers elicit reprogramming at high efficiencies (35–80%), whereas control EVrtTA3 EpiSCs do not yield any iPSCs (Section 4.2.5). Therefore, this is a clean system in which mechanistic study is technically feasible on a reasonable scale.

f. Thanks to the *Rex1::*dGFP reporter, productive reprogramming intermediates can be isolated from the bulk populations (Sections 4.2.2, 4.2.5).

Using this model set, I now return to the key biological questions of the thesis (Section 1.3). In Chapters 5–7, I will interrogate how the naïve pluripotent identity is instructed from EpiSCs by TF/signal interplay, and whether there are multiple routes for this identity transition to occur.

CHAPTER 5:
TRANSCRIPTIONAL TRAJECTORIES ARE DRIVER-SPECIFIC

5.1 Introduction

As described in Chapter 1, a stable cellular identity can be considered as an attractor, occupying a local minimum in an 'energetic' landscape of cell states (Kauffman, 1993; reviewed in Enver et al., 2009). A key outstanding question is whether attractors are multidimensional, with multiple routes of approach to a single identity, or whether a given identity transition follows a single path with a set sequence of mechanistic steps. To answer this fundamental biological question requires a tractable system with well defined start and end identities. Lack thereof has limited progress to date: for example, the landmark work by Huang et al., 2005 demonstrated two distinct routes of promyelocytic HL60 cell differentiation into neutrophils, but they noted disparity in the resulting neutrophil identities and thus had not in fact reached a single attractor state at the end.

Here I seek to understand whether the defined identity transition from EpiSCs to iPSCs can occur via multiple routes. Work of the preceding Chapters has led to establishment, optimisation and validation of an appropriate model set: embryo-derived, hypoxic, FA+XAV-cultured *Rex1::*dGFP EpiSCs with iEsrrb, iPStat3 or iKlf2 single reprogramming drivers. Crucially, these reach the same destination identity (Figure 4.2.8D–E). Now, I interrogate the transcriptional trajectories elicited by each driver to determine whether they follow the same route merely with different speeds/efficiencies, or whether the routes themselves are distinct. By isolation of productive cells based on *Rex1::*dGFP emergence, I will be able to ask whether the mechanistic sequence of events is the same or different within the productive subpopulation of different drivers. For example, in Figure 5.1.1 I illustrate three hypothetical scenarios for reprogramming from EpiSC to iPSC identity. The starred cells represent productive intermediates at an early stage along the reprogramming path(s). Is each starred cell equivalent, i.e. productive cells in scenarios A–C asynchronously follow the same transition route? Or are there differences between the expression signatures of the starred cells, indicative of distinct transition routes?

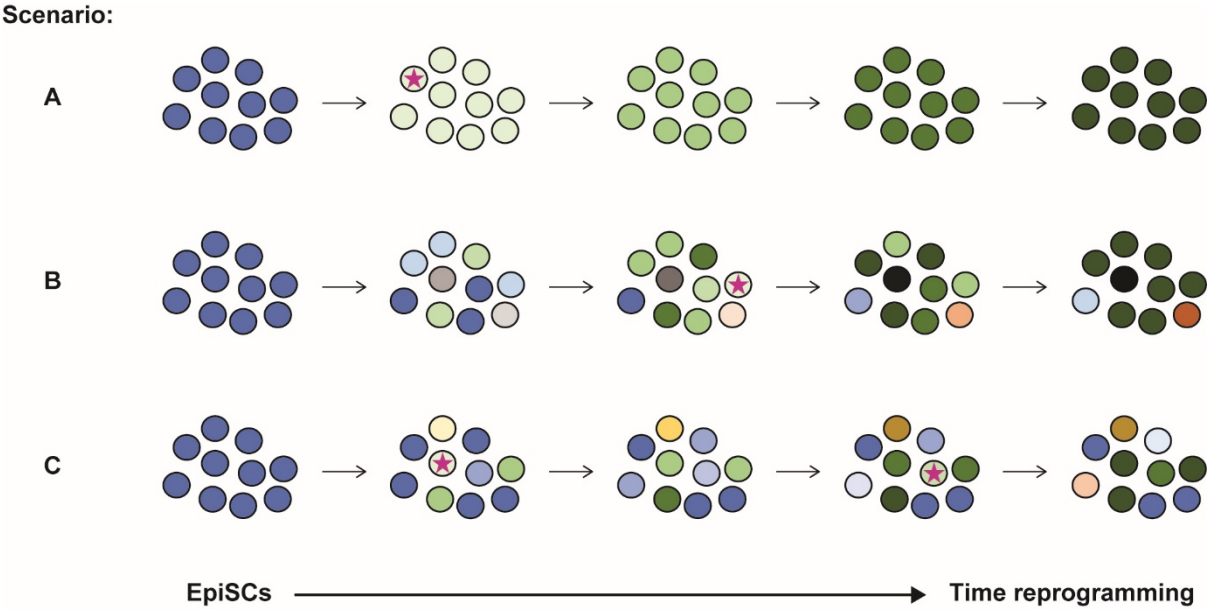


Figure 5.1.1: Are routes distinct or equivalent within the productive subpopulations?

Schematic illustrating the three scenarios of reprogramming. Scenario A is 100% efficient and synchronous. Scenarios B and C more realistically depict asynchronous reprogramming with varying degrees of efficiency and heterogeneity. Blue: EpiSCs. Dark green: iPSCs. Intensifying shades of green: productive progression towards iPSC identity. Starred cells represent productive intermediates at an early stage of reprogramming. Are these equivalent?

Rex1::dGFP reporter activity emerges from day 2 of reprogramming (Section 4.2.5). Therefore, analysis inexorably divides into two temporal phases (Figure 5.1.2). Prior to day 2, there is unfortunately no means of prospectively knowing whether a given cell will reprogram (and it is possible that the cells themselves may not have decided yet). Thanks to the high efficiencies of iEsrrb-, iPStat3- and iKlf2-driven identity transitions (Figure 4.2.8C), it will nevertheless be worth performing analysis of bulk averages at early timepoints in order to ascertain the immediate responses to driver induction (Sections 5.2.1–5.2.2). From day 2 onwards, analysis of single emergent *Rex1*::dGFP+ cells will enable direct comparison of productive reprogramming trajectories (Section 5.2.3).

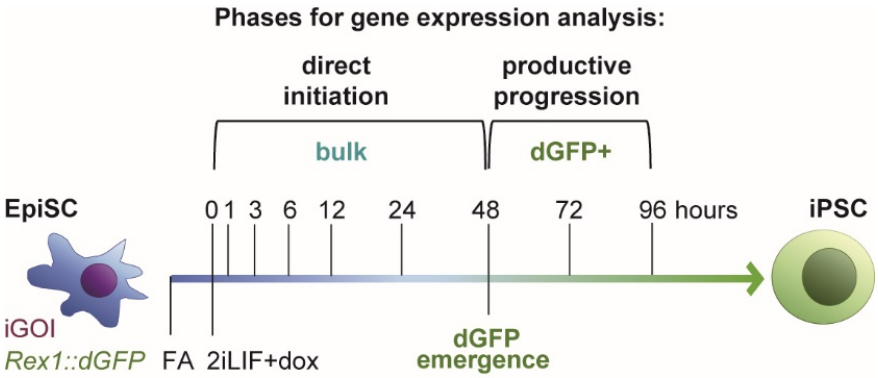


Figure 5.1.2: Overview of reprogramming phases for *Rex1*::dGFP EpiSCs

Schematic of *Rex1::*dGFP EpiSC reprogramming driven by iEsrrb or iKlf2 in 2iLIF+dox, or by iPStat3 in 2iLIF+GCSF. By 48 h, dGFP+ cells are detectable for all lines. Analysis thus divides into two temporal phases. Early initiation events in direct response to transgene induction are assessed in bulk. Following reporter activation, productively progressing cells can be isolated by flow cytometry. iGOI: inducible gene-of-interest.

5.2 Results

5.2.1 Initiation of naïve network wiring is driver-specific

I analysed the initial transcriptional response to each driver from 1–48 hours in bulk (Figures 5.1.2, 5.2.1A). Transgene induction was independently varied with environmental changes, to assess the contribution of and interplay between genetic and signal variables. iPStat3 drove moderate activation of naïve genes in EpiSC conditions, while responses were slightly stronger in 2iLIF, consistent with known boosting of naïve expression by 2i (reviewed in Hackett and Surani, 2014). iEsrrb led to substantially faster and stronger network upregulation in 2iLIF than iPStat3 (Figure 5.2.1A). However, for iEsrrb the response was highly condition-dependent, suggesting that Esrrb and 2iLIF work in synergy to rapidly induce naïve genes.

Surprisingly, Klf2 induction did not lead to upregulation of naïve network genes in any condition, and curiously even silenced Sox2 (Figure 5.2.1A). Considering that Klf2 is a potent reprogramming driver (Figure 4.2.8C), its initial lack of naïve gene induction presented a conundrum. Principle component analysis (PCA) showed a remarkable transcriptional divergence following Klf2 induction (Figure 5.2.1B), and morphological changes during reprogramming initiation were also driver-specific (Figure 5.2.1C). Nonetheless, these divergent routes ultimately re-converged on the same naïve pluripotent destination identity (Figures 4.2.8D–E, 5.2.1B).

I asked which genes could cause such a transcriptional divergence, and found robust upregulation of mesodermal markers in a Klf2-specific manner (Figure 5.2.1D). This indicated initial instigation of a different program downstream of Klf2, rather than simply a delayed naïve induction kinetic. To ascertain whether mesodermal marker expression occurred in cells destined for successful reprogramming or in an unproductive subpopulation, I later turn to single-cell analyses (Section 5.2.3).

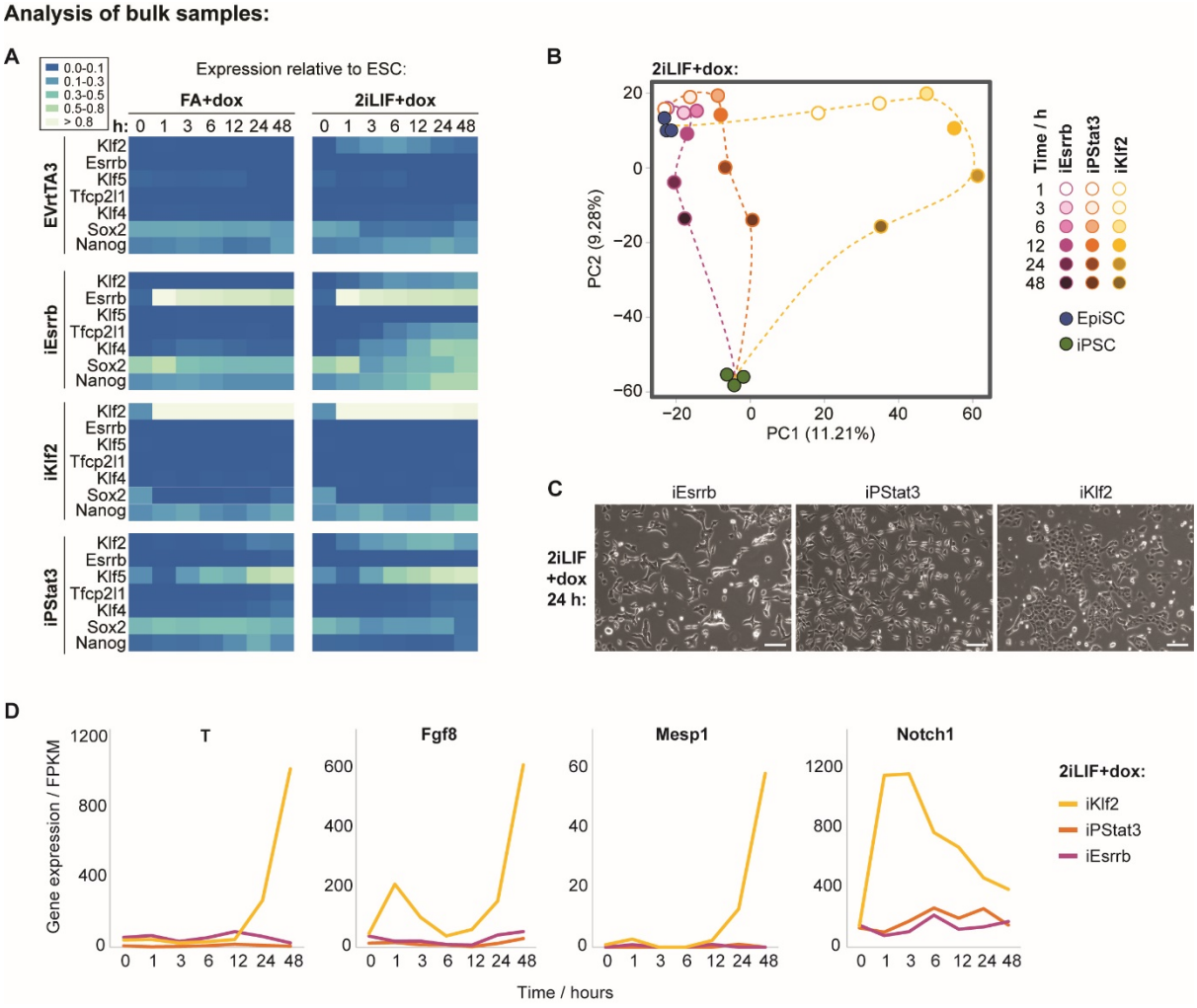


Figure 5.2.1: Initiation of naïve network wiring is driver-specific

(A) Heatmap of mean gene expression from 0–48 h, measured by RT-qPCR relative to Gapdh then normalised to ESC-level=1.
(B) PCA based on the most variable genes (log_2 FPKM>1, CV^2 > 0.5 calculated for each driver then merged to a single list).
(C) Representative phase images 24 h after reprogramming induction. Scale bars: 100 µm.
(D) Gene expression of mesodermal markers following reprogramming induction in 2iLIF.
The RT-qPCR was carried out with technical assistance from Tim Lohoff. The RNAseq analyses were conducted in collaboration with Giuliano Stirparo.

5.2.2 Signal contribution to reprogramming initiation is modulated by the driver

I explored the interaction between drivers and signals in the timecourse bulk RNAseq of reprogramming initiation. *k*-means clustering revealed that iKlf2-driven initiation clustered separately from iEsrrb or iPStat3, irrespective of timepoint or environment (Figure 5.2.2A), corroborating the remarkable divergence of iKlf2-initiation observed in Figure 5.2.1. The signalling environment did not play a strong role in the early transcriptional behaviour of iKlf2-driven reprogramming, with more similarity between timepoints than conditions for iKlf2 (Figure 5.2.2A–B).

In contrast, iEsrrb and 2i components interacted to elicit a transcriptional trajectory distinct from that of iEsrrb in EpiSC FA conditions (Figure 5.2.2A–B). The environmental influence was moderate during iPStat3-driven initiation: 2i signals did not dramatically alter the transcriptional trajectory, but did appear to accelerate it (Figure 5.2.2B).

Together, expression analyses revealed that the pattern and kinetics of naïve network reactivation were driver-dependent, and that signal contribution was markedly modulated by the driver (Figures 5.2.1–2). Nonetheless, these divergent processes ultimately re-converged on the same naïve pluripotent destination identity (Figure 5.2.1C).

Analysis of bulk samples:

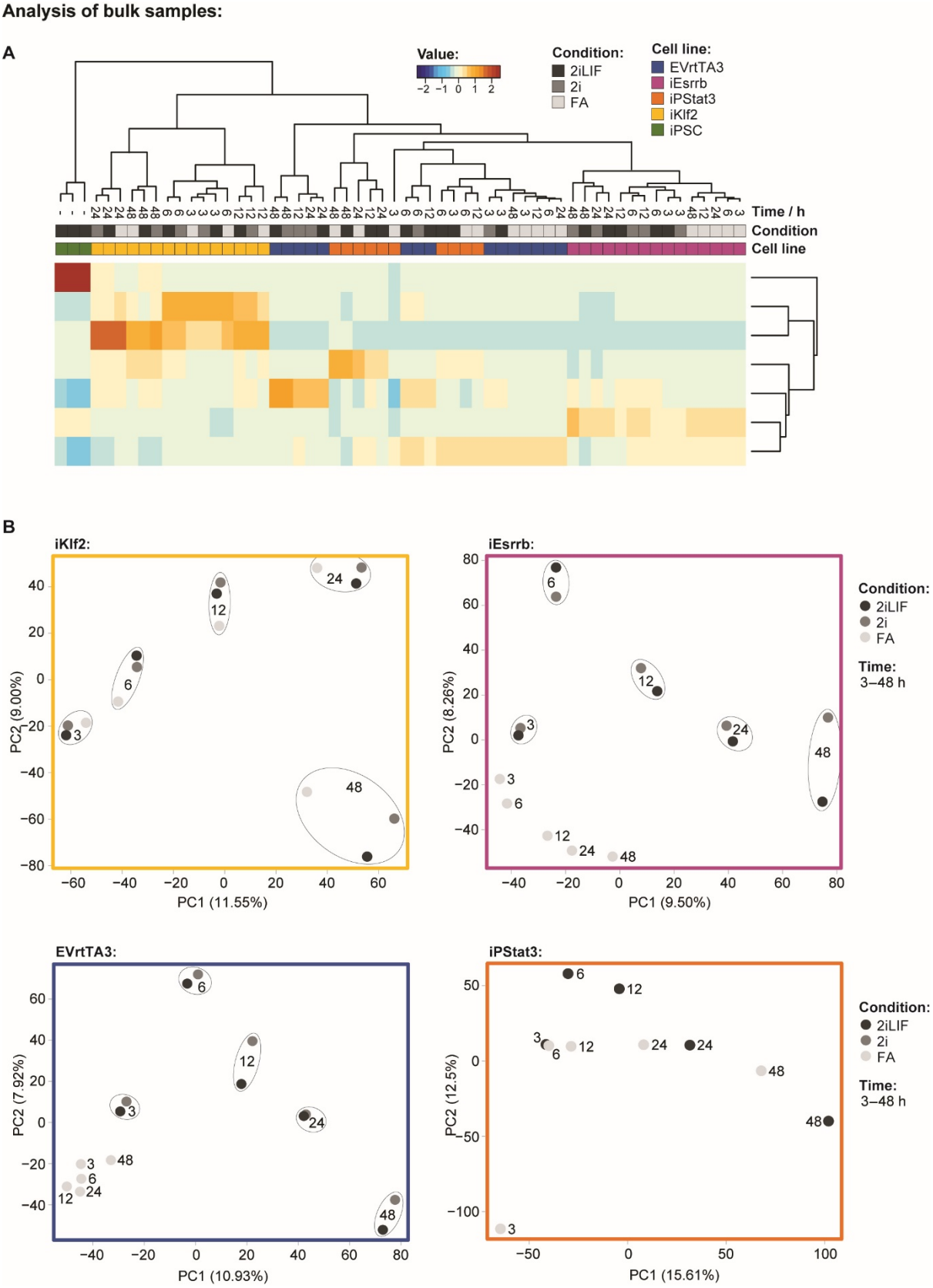


Figure 5.2.2: Signal contribution to reprogramming initiation is modulated by the driver

(A) *k*-means clustering of bulk RNAseq samples, based on expressed genes (FPKM>0).
(B) PCA based on expressed genes (FPKM>0) for each driver.
The RNAseq analyses were conducted in collaboration with Giuliano Stirparo.

5.2.3 Single-cell RNAseq defines distinct productive trajectories

Since reprogramming to naïve pluripotency is heterogeneous and asynchronous, cells undergoing the change of interest must be resolved from the average in order to study transition mechanism(s) (Figures 4.1.1, 5.1.1).

To resolve the logic of the three routes, I performed single-cell (sc)RNAseq on productive intermediates for each driver at 48, 72 and 96 h (Figure 5.2.3A). Intermediate identity was confirmed by naïve vs EpiSC gene expression profiles (Figure 5.2.3B). Sample relationships determined by t-Distributed Stochastic Neighbour Embedding (t-SNE) dimensionality reduction and by hierarchical clustering revealed that *Rex1::*dGFP+ sorted intermediates arranged by driver rather than timepoint (Figure 5.2.3C–D). This demonstrates that reprogramming routes are transcriptionally distinct throughout the productive transitions, not only during bulk initiation. Again, the iKlf2 route was transcriptionally more different from those of iPStat3 and iEsrrb (Figures 5.2.1B, 5.2.3D).

I then examined the kinetics of naïve network activation in single cells. In order to deconvolute the asynchronous nature of reprogramming, cells were ordered by fraction of similarity to origin EpiSCs and destination iPSCs (Figure 5.2.4A). Pseudotime coordinates thus generated largely agreed with real-time. iEsrrb exhibited the fastest kinetics of naïve network orchestration for the majority of naïve genes, while iKlf2 was slowest (Figure 5.2.4B–C). Thus, single-cell analyses confirmed the different kinetics of naïve network activation observed in bulk initiation (Figure 5.2.1A) and extended these to later timepoints.

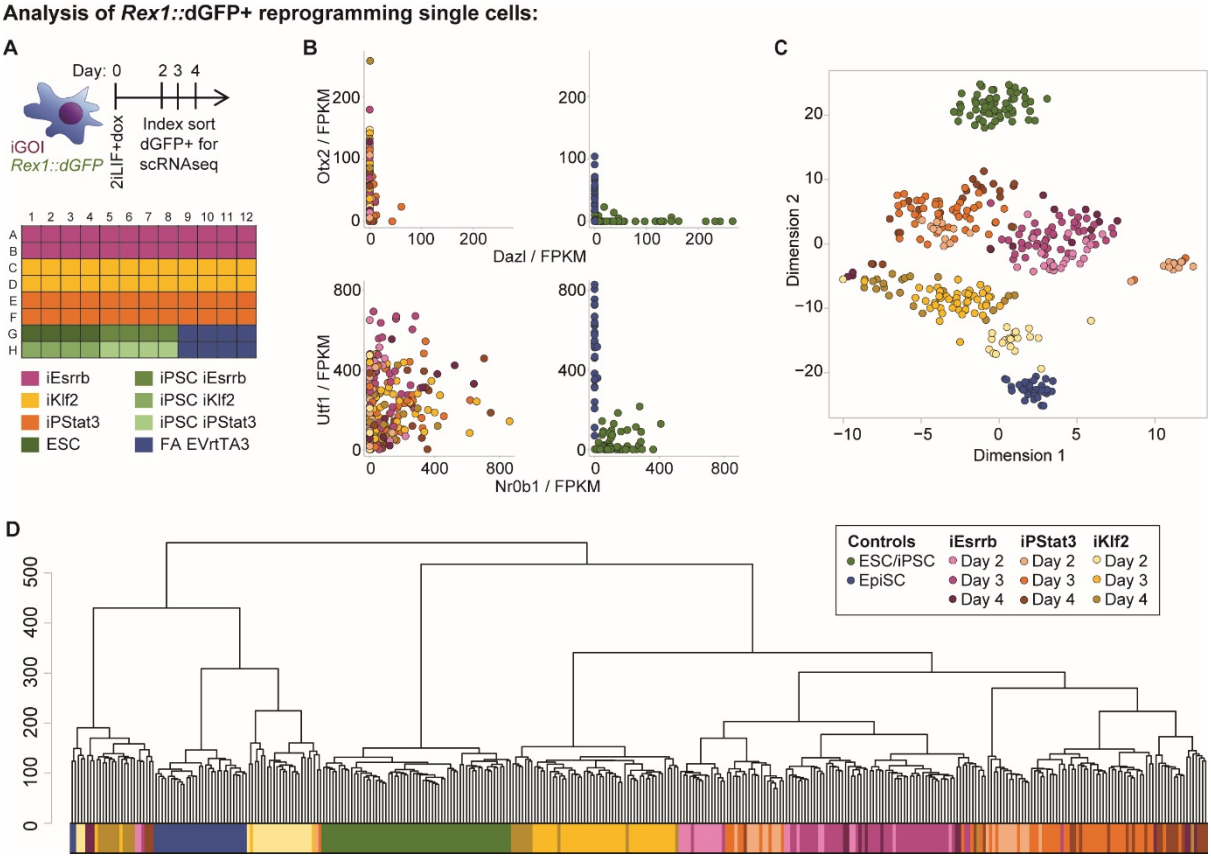


Figure 5.2.3: Single-cell RNAseq of productive reprogramming intermediates

(A) Schematics depicting the isolation of productive single cells from each driver line, and the 96-well plate layout into which single cells were index-sorted for scRNAseq. One, two and one such plates were prepared for days 2, 3 and 4 respectively. Cells from all lines were included on each plate on each day, to minimise downstream batch effects. Sorted reprogramming intermediates (2iLIF+dox/GCSF) and ESCs/iPSCs (2iLIF) were *Rex1*::dGFP+, whereas EpiSCs (FA) were *Rex1*::dGFP-. iGOI: inducible gene-of-interest.
(B–D) Shared legend.
(B) Expression scatter plots of genes expressed in EpiSCs (Otx2, Utf1) or in naïve ESCs/iPSCs (Dazl, Nr0b1).
(C) t-SNE plot showing relationships between sequenced single cells.
(D) Unsupervised hierarchical cluster of sequenced single cells, computed with the Ward.D2 agglomeration method and Euclidean distances.
The RNAseq analyses were conducted in collaboration with Giuliano Stirparo.

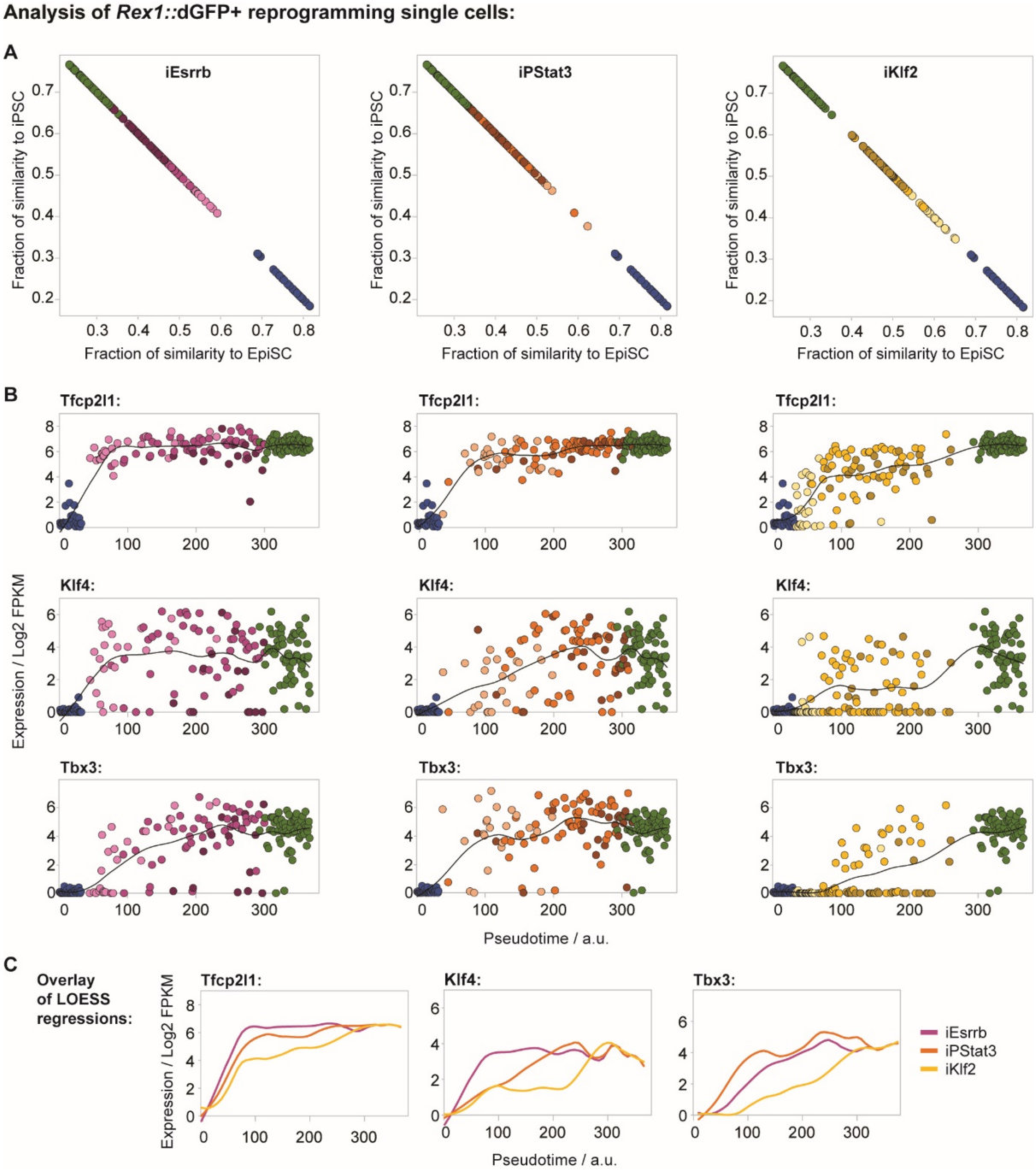


Figure 5.2.4: Kinetics of naïve network induction in productive single cells

(A) Computation of fraction of identity between each single cell vs EpiSCs in FA and iPSCs in 2iLIF. Signature EpiSC and iPSC datasets were generated by averaging of bulk RNAseq samples. Single cells were ordered from lowest to highest iPSC/EpiSC identity fractions to generate pseudotime coordinates from reprogramming start to end.
(B) Scatter plots of expression vs pseudotime, fitted with LOESS regression lines.
(C) Overlay of LOESS regression fit lines from the above plots, to facilitate comparison.
The RNAseq analyses were conducted in collaboration with Giuliano Stirparo.

Strikingly, the induction of mesodermal markers revealed during bulk initiation persisted in productive single cells for iKfl2 (Figures 5.2.1D, 5.2.5A–B). This indicated that transient activation of mesodermal markers was not just due to differentiation of a population of unproductive cells, but was a transcriptional response occurring during productive[3] establishment of naïve pluripotency when driven by Klf2.

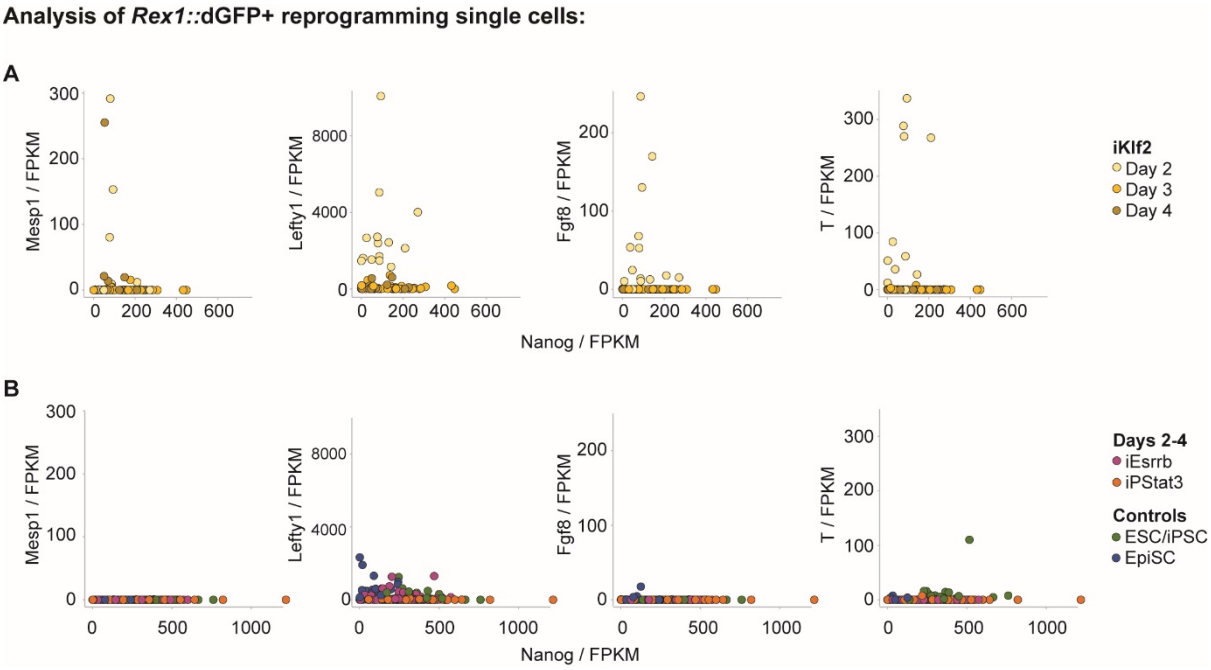


Figure 5.2.5: Expression of mesodermal markers in single iKlf2 intermediates

(A) Scatter plots of mesodermal markers (Mesp1, Lefty1, Fgf8, T) against Nanog expression in single iKlf2 reprogramming intermediates.
(B) Scatter plots of mesodermal markers against Nanog expression in single iEsrrb and iPStat3 intermediates, EpiSC (start) and ESC/iPSC (end) identity controls. The *y*-axes are to the same scales as above to facilitate comparison.

[3] Since first submission of this thesis, I generated *T/Rex1* double-reporter EpiSCs and performed reporter live-imaging of iKlf2-driven reprogramming. This confirmed that T+ intermediates convert into Rex1+ iPSCs, i.e. that productive iKlf2 reprogramming proceeds via a T+ state on the protein level. Sequential activation of T then Rex1 reporter is consistent with the low % of T+ cells captured by scRNAseq of Rex1+ intermediates in Figure 5.2.5A above. For details, please see Appendix page 178.

To place the reprogramming trajectories in the context of development, we compared scRNAseq of productive intermediates with E2.5–E6.5 embryos (Deng et al., 2014; Mohammed et al., 2017). Single-cell transcriptome analyses revealed that iPStat3 reprogramming intermediates transiently acquired more similarity to the early inner cell mass (ICM) (Figure 5.2.6A–B) and exhibited a Nanog+Gata6+ double-positive signature (Figure 5.2.7A–B). Nanog+Gata6+ co-expression arises in the early ICM of pre-implantation embryos (Figure 1.1.1A) (Plusa et al., 2008), prompting the hypothesis that iPStat3-driven reprogramming goes further back to an early ICM-like state, then forwards into the consolidated naïve identity. There is a temporal sequence of naïve gene activation in the embryo (Boroviak et al., 2015). Earlier markers (Klf2, Tfcp2l1) are expressed in Nanog+Gata6+ early ICM. In contrast, later markers (Gbx2, Nr0b1) are not activated until after Nanog+Gata6- naïve epiblast has segregated from Nanog-Gata6+ primitive endoderm (PrE). iPStat3 reprogramming intermediates emulated this *in vivo* progression: Klf2 and Tfcp2l1 were turned on earlier and are co-expressed with Gata6, whereas Gbx2 and Nr0b1 were activated later in Gata6- cells (Figure 5.2.7C). In sum, iPStat3-driven reprogramming recapitulated defining molecular features of naïve pluripotency establishment *in vivo.* In contrast, iKlf2 intermediates retained similarity to later developmental stages and transiently expressed mesodermal lineage markers (Figures 5.2.5A, 5.2.6B).

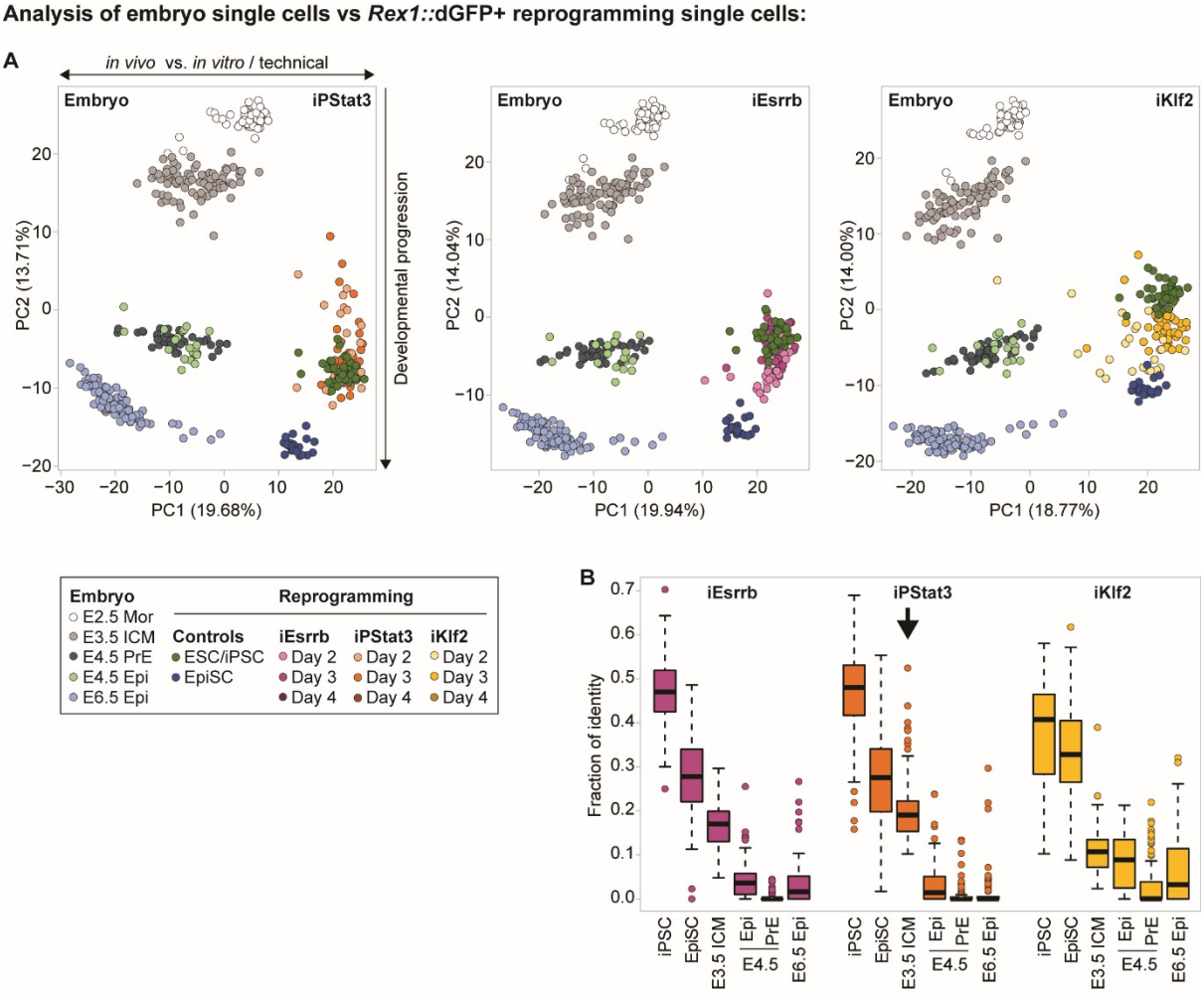


Figure 5.2.6: Developmental context of productive reprogramming intermediates

(A) PCA based on the most variable genes (log_2 FPKM>1, CV^2>0.5) for reprogramming intermediates and embryo single cells. Mor: compacted morula. ICM: inner cell mass. Epi: epiblast. PrE: primitive endoderm. PC1 separates *in vivo* vs *in vitro* datasets; PC2 portrays developmental progression[4].
(B) Fraction of similarity to signature datasets[5] was computed by quadratic programming for each reprogramming single cell (days 2, 3, 4), and is presented as box-and-whisker plots.
The RNAseq analyses were conducted in collaboration with Giuliano Stirparo.

[4] Please note that since the sorted reprogramming cells were already Rex1+, early stages of greater diversion are not represented here, particularly that of Klf2 towards mesoderm. Subsequent sequencing that captures this can be found in the Appendix page 176.

[5] For analysis of day 2 only with an improved design, please see Appendix page 179.

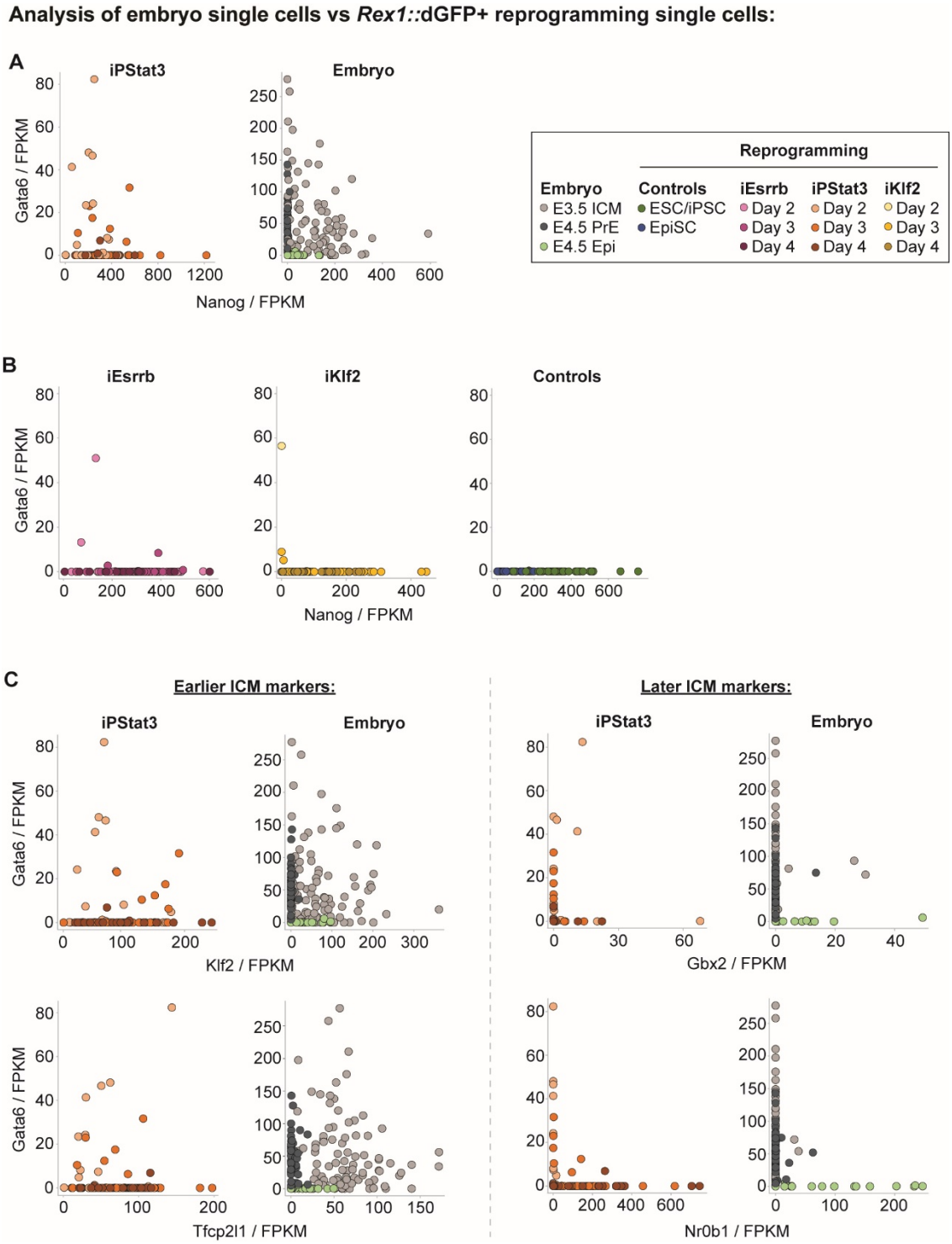


Figure 5.2.7: Nanog+Gata6+ iPStat3 intermediates emulate developmental progression

(A–B) Scatter plots of Gata6 vs Nanog expression in iPStat3, iEsrrb and iKlf2 reprogramming intermediates, E3.5 and E4.5 embryos. EpiSC (start) and ESC/iPSC (end) identity controls. ICM: inner cell mass. Epi: epiblast. PrE: primitive endoderm.
(C) Scatter plots of Gata6 vs earlier or later ICM markers in iPStat3 reprogramming intermediates, E3.5 and E4.5 embryos.

5.3 Discussion

The work of Chapter 5 demonstrates that iEsrrb, iPStat3 and iKlf2 each drive identity transition from EpiSC to naïve iPSC identity along a distinct transcriptional trajectory (Figure 5.3.1). This provides empirical evidence supporting theories of cellular identity as a multidimensional attractor (Huang et al., 2005; Kauffman, 1993), whereby there are multiple routes of approach towards a single identity. In the future, I hope that relevant experts will formally model the RNAseq data of Chapter 5 using dynamical systems mathematics.

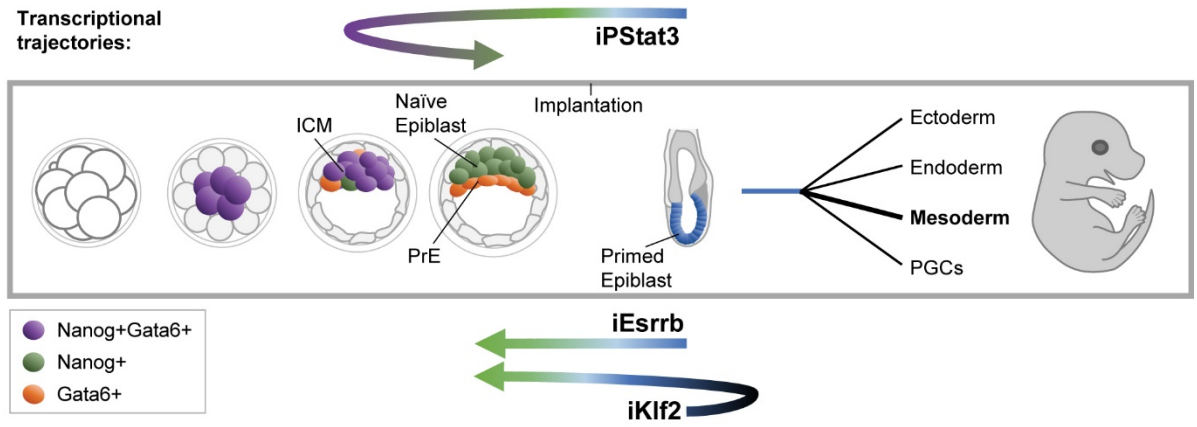


Figure 5.3.1: Summary of transcriptional trajectories for reprogramming routes

Schematic of reprogramming transcriptional trajectories relative to early development.

Returning to the hypothetical scenarios outlined in Figure 5.1.1, this work shows that the starred cells are not equivalent. In addition to differences in reprogramming kinetics and efficiencies, the single productive intermediates for each driver have distinct transcriptional signatures (Section 5.2.3). Furthermore, the influence of signals on the transcriptional trajectory was driver-dependent (Figures 5.2.1–2), indicative of distinct underlying transition mechanisms. This theme is explored further in Chapter 6.

The rapid upregulation of naïve genes by iEsrrb is intuitive for an efficient reprogramming driver. In contrast, iKlf2 presents a fascinating conundrum. Not only does it fail to upregulate naïve genes within the first 48 h (Figure 5.2.1A), it follows a highly divergent initiation trajectory (Figure 5.2.1B), exhibits mesodermal marker expression even in productive intermediates (Figure 5.2.5A) and clusters separately from other drivers at all times and in all conditions (Figures 5.2.2A, 5.2.3C–D). How, then, does this ultimately provide the logic for an efficient transition into the naïve pluripotent identity? I return to this paradox in Chapter 7.

iPStat3-driven reprogramming transcriptionally recapitulates key aspects of naïve epiblast specification *in vivo* (Figures 5.2.6–7). This suggests that, in some cases, reprogramming can overshoot backwards in development then proceed forwards again into the consolidated naïve identity. I wonder whether this action of PStat3 in reprogramming reflects its role(s) during *in vivo* pre-implantation development. PStat3 expression is observed *in vivo* from the 8-cell stage onwards, and it has recently been reported to play a context-dependent role in PrE as well as epiblast fate specification (Anderson et al., 2017; Morgani and Brickman, 2015). If transferred to an appropriate environment, could the PStat3-driven Nanog+Gata6+ reprogramming intermediates be open to PrE as well as epiblast fates[6]?

[6] Since first submission of this thesis, I performed this experiment and found that iPStat3 Gata6+ intermediates do indeed contribute to PrE as well as epiblast. Please see Appendix page 179.

CHAPTER 6:
ROUTES ARE MECHANISTICALLY DISTINCT

6.1 Introduction

Demonstration of transcriptionally distinct routes for identity transition between EpiSCs and iPSCs provided evidence in support of cellular identity as a multidimensional attractor (Chapter 5). Some degree of transcriptional divergence might be expected simply due to different direct gene targets of each TF driver, although this is unlikely to offer a sufficient explanation given the extent of differences when placed in a developmental context (Figures 5.2.5–5.2.7). To conclusively determine whether transition routes are distinct, mechanistic attributes must also be considered.

Therefore, in Chapter 6 I independently manipulate genetic and environmental parameters, interrogating whether the requirements for and interplays between TFs and signals are driver-dependent during reprogramming of EpiSCs to naïve pluripotency. I then test my findings in other contexts of naïve pluripotency establishment: development of the naïve epiblast in mouse embryos, and resetting of primed human ESCs towards naïve pluripotency.

6.2 Results

6.2.1 Downstream genetic mediators of reprogramming drivers

To test whether the aforementioned transcriptional trajectories were indicative of differences in their transition mechanisms, I assessed the routes' genetic and signal requirements. Putative downstream genetic mediators were identified by examining expression of known reprogramming drivers 24 hours after induction of iEsrrb, iKlf2 or iPStat3 (Figure 6.2.1A). Endogenous Esrrb was not strongly upregulated by either iKlf2 or iPStat3 by 24 hours, and is thus unlikely to be implicated in their reprogramming initiation mechanisms. In contrast, endogenous Klf2 reached 50% and 20% of ESC-level in iPStat3 and iEsrrb respectively.

Given this early response, I tested whether Klf2 was a functional mediator of iPStat3- or iEsrrb-driven reprogramming by knock-down (KD) of Klf2 at reprogramming onset. I found that transient Klf2 KD abolished reprogramming driven by iPStat3 but not iEsrrb (Figure 6.2.1B–D). This implicated Klf2 as a critical functional mediator of reprogramming initiation by iPStat3. Curiously, iPStat3 reprogramming sensitivity to Klf2 KD was somewhat context-dependent, partially alleviated in the absence of PD03 (Figure 6.2.1E).

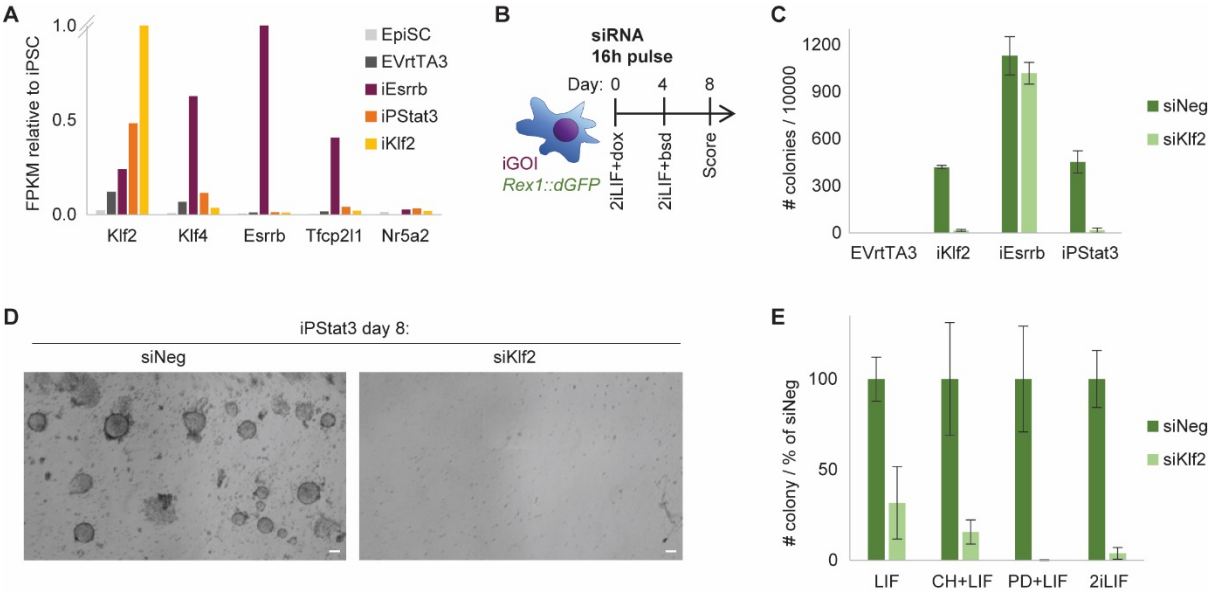


Figure 6.2.1: Route-specific dependence on Klf2 as a reprogramming mediator

(A) Gene expression of selected reprogramming drivers 24 h after induction of iEsrrb, iKlf2, iPStat3 or EVrtTA3 in 2iLIF+dox/GCSF, relative to iPSC-level=1. EpiSC control is in FA. *y*-axis iEsrrb: Esrrb=3.32. iKlf2: Klf2=8.30.
(B–D) Klf2 KD was performed at reprogramming onset with a single pulse of siRNA. iGOI: inducible gene-of-interest. iPSC colonies were scored on day 8, presented as mean ±SD (n=3). Representative phase images are shown on day 8. Scale bars: 100 µm.
(E) Klf2 KD was performed at iPStat3-driven reprogramming onset with a single pulse of siRNA, in the indicated conditions +GCSF from day 0–4, then iPSC colonies were selected from day 4–8 in 2iLIF+blasticidin. iPSC colonies were scored on day 8, presented as mean ±SD (n=3). PD=PD03; CH=Chiron; 2i=PD+CH.

6.2.2 Exogenous signal requirements for each reprogramming route

To assess mechanistic differences between routes in terms of exogenous signal requirements, I challenged the first 4 days of reprogramming with different 2iLIF signal permutations (Figure 6.2.2A–B). iPStat3 yielded iPSCs in the absence of both PD03 and Chiron, but together PD03 and Chiron synergistically boosted the efficiency. However, the impact of PD03 and Chiron was driver-dependent: Chiron was essential for iKlf2-driven reprogramming, with no benefit from additional supplementation with PD03. Functional redundancy between Klf2 and PD03 has been previously noted (Yeo et al., 2014), and the inability of iKlf2 to drive reprogramming without direction from an exogenous signal is in agreement with the observation that iKlf2 does not directly induce naïve gene expression (Figure 5.2.1A). Unlike iKlf2, reprogramming driven by iEsrrb was highly LIF-dependent (Figures 6.2.2B, 6.2.3A).

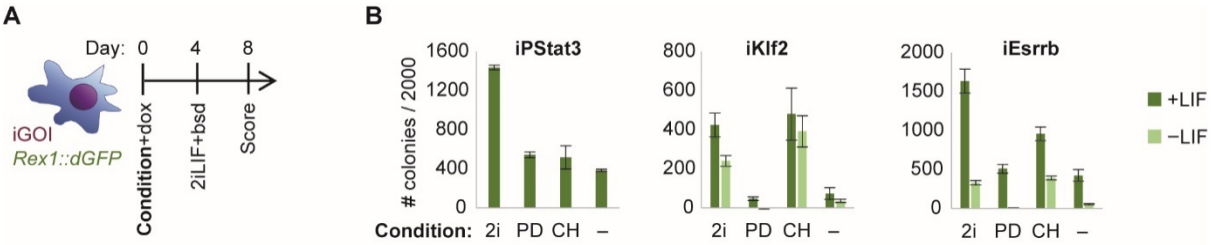


Figure 6.2.2: Route-specific requirements for exogenous signals

(A) Schematic depicting the experimental strategy. Reprogramming was induced in different conditions from day 0–4, then iPSC colonies were selected from day 4–8 in 2iLIF+blasticidin (bsd). iPSC colonies were scored on day 8.
(B) Mean number of iPSC colonies on day 8, ±SD (n=3). PD=PD03; CH=Chiron; 2i=PD+CH.

6.2.3 Differential modulation of exogenous signal transduction to target genes

iEsrrb induction in LIF led to greater upregulation of canonical PStat3 targets than induction of iPStat3 itself (Figure 6.2.3B).This was not due to elevation of PStat3 protein by Esrrb (Figure 6.2.3C) and thus demonstrated downstream synergy between Esrrb and PStat3 that amplified LIF/Stat3 transduction to the naïve TF network.

A turning point occurred 6 h after induction: from 0–6 h, Klf2 was upregulated similarly in 2i±LIF for both iEsrrb and negative control EpiSCs; after 6 h, Klf2 expression continued to increase in iEsrrb+2iLIF, but collapsed in iEsrrb+2i and all control conditions (Figure 6.2.3D–E). Klf4 upregulation also launched in earnest after 6 h with iEsrrb+2iLIF. Thus, cooperation between Esrrb and LIF became effectual after 6 h, consolidating and extending the 2i-mediated instigation of naïve network orchestration.

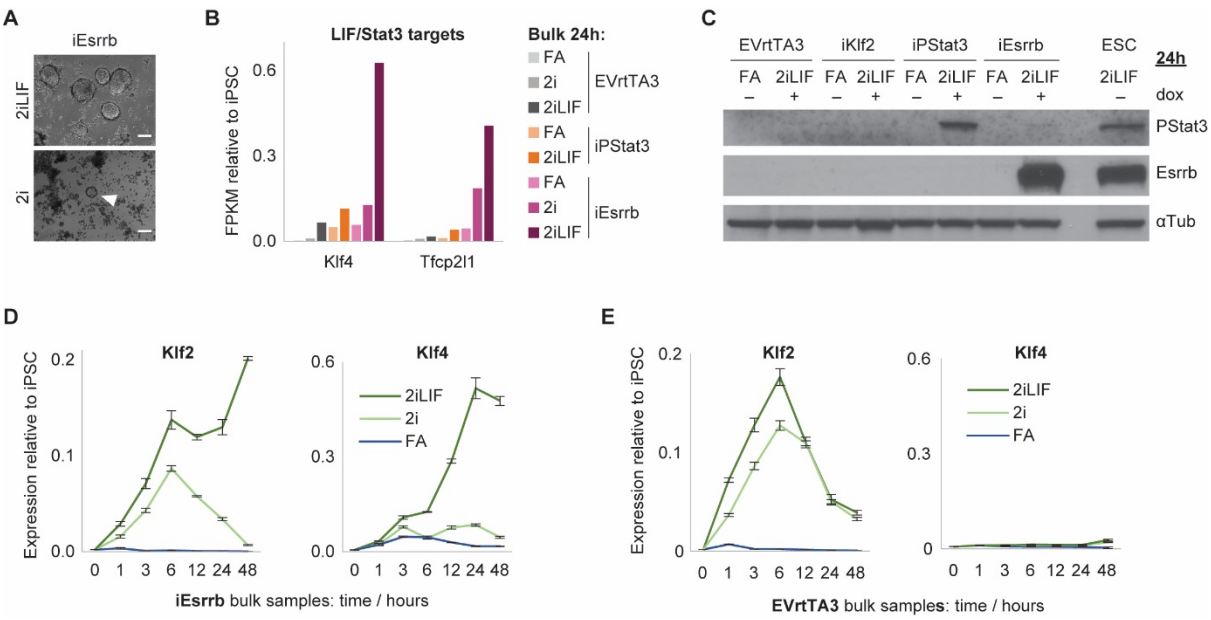


Figure 6.2.3: Synergy with Esrrb amplifies LIF/Stat3 target gene upregulation

(A) Representative phase images of iEsrrb on day 8 in 2iLIF+blasticidin, after reprogramming from day 0–4 in 2iLIF+dox or 2i+dox as indicated. White arrow indicates an iPSC colony. Scale bars: 100 μm.
(B) Gene expression of selected LIF/Stat3 target genes (Martello et al., 2013; Niwa et al., 2009; van Oosten et al., 2012; Ye et al., 2013) 24 h after induction of iEsrrb, iPStat3 or EVrtTA3 in the indicated conditions, relative to iPSC-level=1.
(C) Western blot against PStat3, Esrrb and αTubulin, after 24 h in FA or 2iLIF+dox/GCSF.
(D–E) RT-qPCR analysis of iEsrrb **(D)** or EVrtTA3 **(E)** EpiSCs in the indicated condition +dox. Mean gene expression is displayed relative to Gapdh and normalised to iPSC-level=1, ±SD (n=3). The iEsrrb/EVrtTA3 axes are to the same scales to facilitate comparison.

6.2.4 Identification of endogenous BMP signalling as a route-specific mechanism

In light of the above observations that signal requirements and interpretation are driver-dependent, I interrogated RNAseq data of bulk initiation and single dGFP+ reprogramming intermediates for evidence of other signalling differences between iKlf2, iEsrrb and iPStat3. BMP signalling pathway target Id1 was upregulated in iKlf2 and iPStat3, but not iEsrrb (Figure 6.2.4A–B). Id1 upregulation was intermediate-specific, with negligible expression in starting EpiSCs or destination iPSCs. BMP signalling is the key pluripotency-sustaining component in the serum of classical ESC cultures (Ying et al., 2003), is important for MET in serum-based fibroblast reprogramming (Chen et al., 2011; Samavarchi-Tehrani et al., 2010), but is not thought to be active in 2iLIF-cultured ESCs (Boroviak et al., 2014).

I assessed BMP pathway status by PSmad1/5 immunofluorescence during EpiSC reprogramming in 2iLIF, finding positive staining for iKlf2 and iPStat3 but not iEsrrb (Figure 6.2.4C). Therefore, BMP signalling is activated in a route-specific manner. Both the speed (Figure 6.2.4A) and strength (Figure 6.2.4C) of BMP signalling activation were greater for iKlf2 than iPStat3. For iKlf2, Id1 was sharply upregulated at 1 hour, substantially prior to naïve gene expression upregulation, whereas for iPStat3 the two were more concomitant (Figure 5.2.1A). These differences in signalling dynamics represent additional driver-specific features of interest.

To test whether auto/paracrine BMP signalling is required during EpiSC reprogramming, I applied DMH2 from day 0–4. DMH2 is a specific and well characterised BMP signalling inhibitor (Figure 6.2.4D) (Hao et al., 2010). DMH2 treatment abolished iKlf2- and iPStat3-driven reprogramming in 2iLIF, but iPSC colonies still formed for iEsrrb (Figure 6.2.4E–G). Therefore, DMH2 treatment blocked reprogramming only in those lines exhibiting evidence of active BMP signalling in their intermediates. This was specific to the transition, being dispensable for maintenance of resultant iPSCs in 2iLIF (Figure 6.2.5A–B).

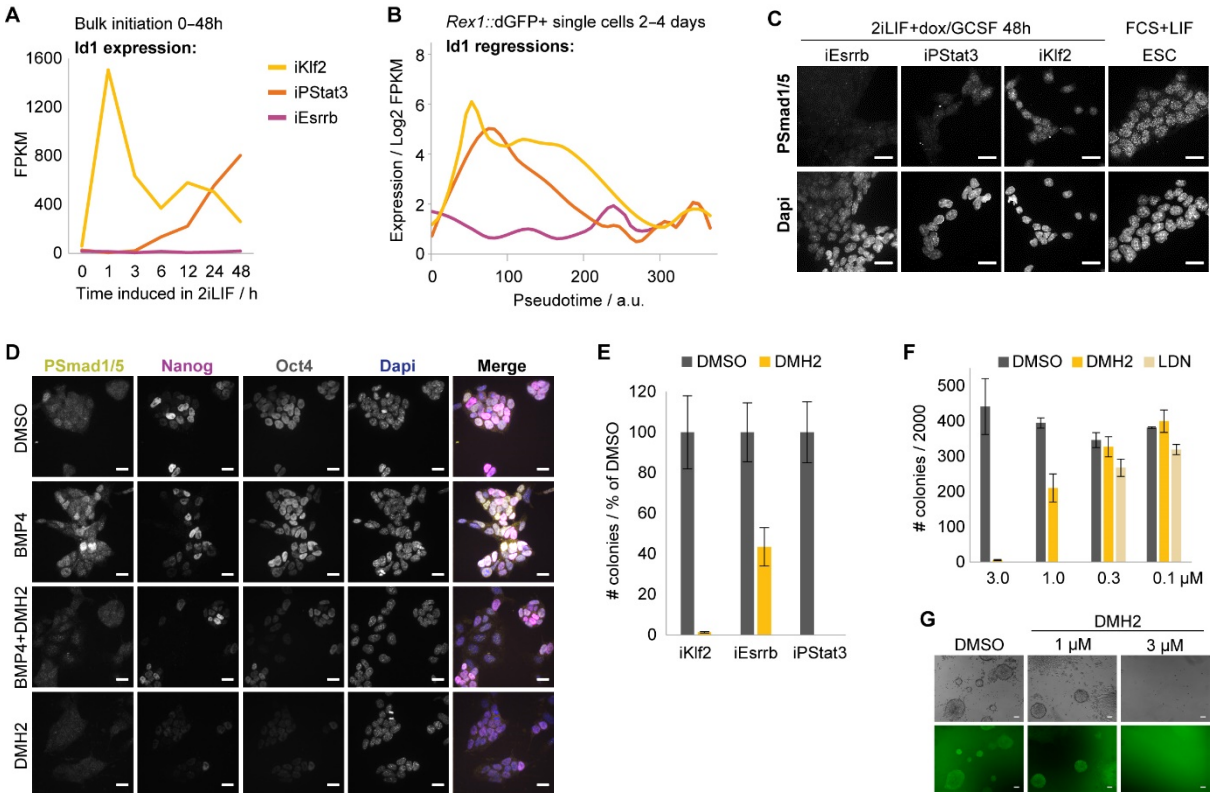


Figure 6.2.4: BMP signalling requirements for naïve pluripotency specification

(A) Id1 expression in bulk RNAseq following reprogramming initiation in 2iLIF.
(B) LOESS regression fit lines summarise Id1 kinetics during reprogramming, computed from scatter plots of $\log_2$ FPKM vs pseudotime for single cells (Fig. S2D) by Giuliano Stirparo.
(C) Immunofluorescent staining 48 h after reprogramming induction. ESCs cultured in FCS+LIF provide a positive control for PSmad1/5. Representative maximum intensity projections of Z-stack 60x slices are presented. Scale bars: 20 μm. Technical assistance was provided by Andrew Malcolm.
(D) Validation of 3 μM DMH2 efficacy. Wild-type E14 ESCs cultured in FCS+LIF were treated for 24 h as indicated. BMP signalling is known to be active and important for pluripotency maintenance in FCS+LIF (Ying et al., 2003), unlike in 2iLIF. Representative maximum intensity projections of Z-stack 40x slices are presented. Scale bars: 20 μm.
(E) 3 μM DMH2 or 1:1000 DMSO was applied to reprogramming in 2iLIF+dox/GCSF from day 0–4, then iPSC colonies were selected in 2iLIF+blasticidin from day 4–8. iPSC colonies were scored on day 8, presented as mean ±SD (n=3).
(F–G) Titration of DMH2 and LDN concentrations, applied as indicated during day 0–4 of iKlf2-driven reprogramming in 2iLIF+dox. iPSC colonies were selected in 2iLIF+blasticidin from day 4–8, scored on day 8 and presented as mean ±SD (n=3). Representative day 8 phase and dGFP images are shown. Scale bars: 100 μm. LDN is a different BMP signalling inhibitor (Cuny et al., 2008), which we used to independently verify DMH2 phenotypes.

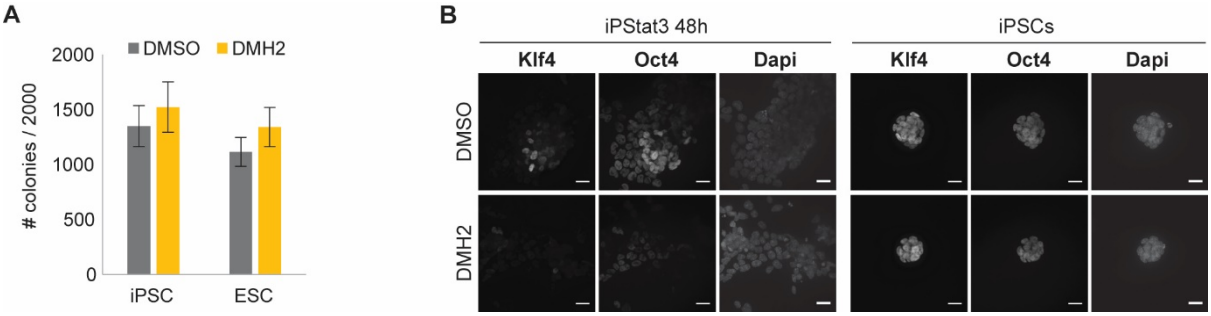


Figure 6.2.5: Endogenous BMP signalling is dispensable for resultant iPSCs

(A) 3 μM DMH2 or 1:1000 DMSO was applied to ESCs or previously established iKlf2 iPSCs, at clonal density in 2iLIF. Naïve colonies were scored after 4 days, presented as mean ±SD (n=3).
(B) Representative immunofluorescent staining of Klf4 and Oct4 expression after 48 h of treatment with 3 μM DMH2 or 1:1000 DMSO, for EpiSC reprogramming driven by iPStat3 in 2iLIF+GCSF, or for previously established iPStat3 iPSCs in 2iLIF. Scale bars: 20 μm.

6.2.5 BMP signalling is required for naïve pluripotency establishment *in vivo* and during human resetting

Having identified BMP signalling requirement in two routes of mouse EpiSC reprogramming, and given that iPStat3 reprogramming intermediates transiently acquired similarity to the early ICM, I explored whether endogenous BMP signalling also plays a role in other contexts of naïve pluripotency establishment: development of the naïve epiblast in mouse embryos, and resetting of primed human ESCs towards naïve pluripotency.

Conventional human ESCs (hESCs) exhibit primed pluripotency, like mouse EpiSCs (reviewed in Davidson et al., 2015). Recently, substantial progress has been made to convert primed hESCs to the putative human naïve pluripotent state (Guo et al., 2017; Takashima et al., 2014; Theunissen et al., 2014), but the transcriptional logic and signalling cues governing this process remain poorly understood. Given that one method of hESC resetting is driven by inducible expression of Klf2+Nanog (iK2N), and that Klf2 upregulates BMP signalling in mouse EpiSCs (Figure 6.2.4A–C), I tested whether endogenous BMP signalling also plays a role in hESC resetting.

DMH2 treatment led to a dramatic reduction in colony size after 7 days of induction with 2iLIF+dox (Figure 6.2.6A–C), and no colonies remained after 1 passage (Figure 6.2.6D). Prior to culture collapse, DMH2-treated hESCs exhibited a reduction in naïve network expression (Figure 6.2.6E). Therefore, I propose that BMP signalling is required for iK2N-mediated resetting of primed hESCs towards naïve pluripotency.

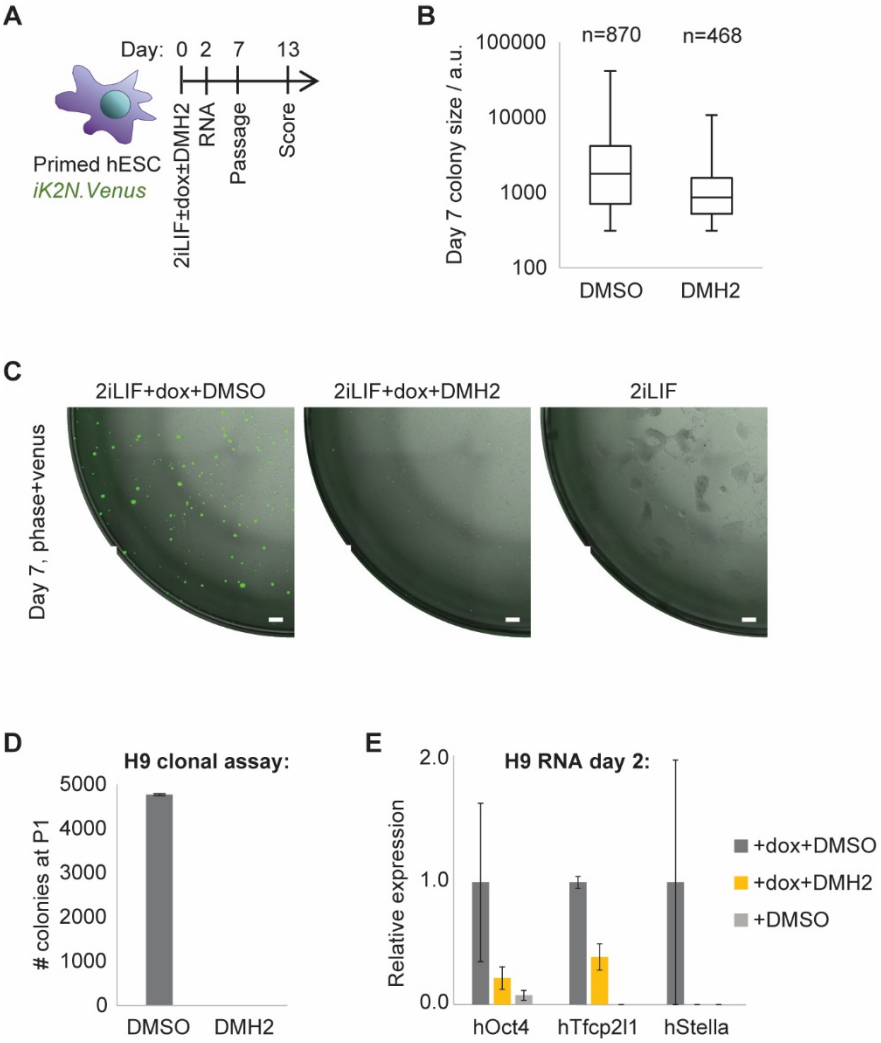


Figure 6.2.6: Effect of DMH2 treatment on resetting of primed human ESCs

(A) Resetting towards the naïve identity was initiated for human H9 *iK2N.Venus* primed ESCs by application of 2iLIF+dox ± 3 μM DMH2.
(B–C) Size of Venus+ colonies was measured on day 7 of H9 iK2N.Venus resetting in 2iLIF+dox ± 3 μM DMH2. Box-and-whisker plots are displayed on a $\log_{10}$ scale **(B)**, together with representative images of merged phase and Venus channels **(C)**. Venus+ distinguishes resetting H9 cells from feeders. Scale bars: 500 μm.
(D) Mean number of Venus+ P1 colonies on day 13 is presented ±SD (n=3). Venus+ distinguishes resetting H9 cells from feeders.
(E) RT-qPCR analysis of H9 *iK2N.Venus* after 2 days of treatment in the indicated conditions, which is prior to culture collapse with DMH2. Mean gene expression is displayed relative to Gapdh and normalised to the highest level for each gene.

The BMP signalling pathway is active in pre-implantation mouse embryos from the 4-cell stage onwards, including in the ICM (Graham et al., 2014; Reyes de Mochel et al., 2015), so involvement in epiblast specification is plausible. I cultured embryos ±DMH2 from the late morula stage, then assessed naïve network status in the mid–late blastocyst by immunofluorescence (Figure 6.2.7A–C). The ICM still formed in DMH2-treated embryos, indicated morphologically and by the cluster of Oct4-positive cells. However, expression of Nanog and Klf4 was dramatically reduced. Thus, BMP signalling inhibition applied from the late morula disrupted the induction of embryonic naïve pluripotency[7]. Identification of this novel role for the BMP pathway *in vivo* highlighted the power of defined reprogramming systems to uncover principles of identity specification.

These observations were somewhat at odds with the work of Graham et al., 2014, in which they did not find differences in Nanog expression following dorsomorphin treatment but instead reported defects in extra-embryonic lineage development. However, dorsomorphin is not a specific BMP inhibitor, but is primarily an AMPK inhibitor (Zhou et al., 2001) and also inhibits VEGF signalling (Hao et al., 2010). In contrast, DMH2 selectively inhibits ALK6, ALK3 and ALK2 (Hao et al., 2010). It is also possible that disparities stemmed from mouse strain differences: in the work of Graham et al., 2014, embryos were harvested from superovulated C57Bl6xCBA mice, whereas my investigation was performed from normal matings of strain 129 mice.

Furthermore, I noted that the epiblast phenotype was highly sensitive to the timing of BMP inhibition (Figure 6.2.7D). Addition of DMH2 at the 8-cell stage caused developmental arrest, consistent with Reyes de Mochel et al., 2015, but precluding fair assessment of whether the naïve lineage specifically was compromised: at the 8-cell stage, the trophectoderm vs ICM decision has not yet been made. Conversely, application of DMH2 to the mid blastocyst did not disrupt the naïve epiblast. By this point, the epiblast vs PrE bifurcation is already underway. In contrast, precisely timed addition of DMH2 at cavitation onset in the late morula falls between these two developmental lineage bifurcations, and thus permitted assessment of the role of BMP signalling in naïve epiblast specification despite the multitude of BMP signalling roles during early development.

[7] For an extended analysis performed since first submission of this thesis, see Appendix page 181.

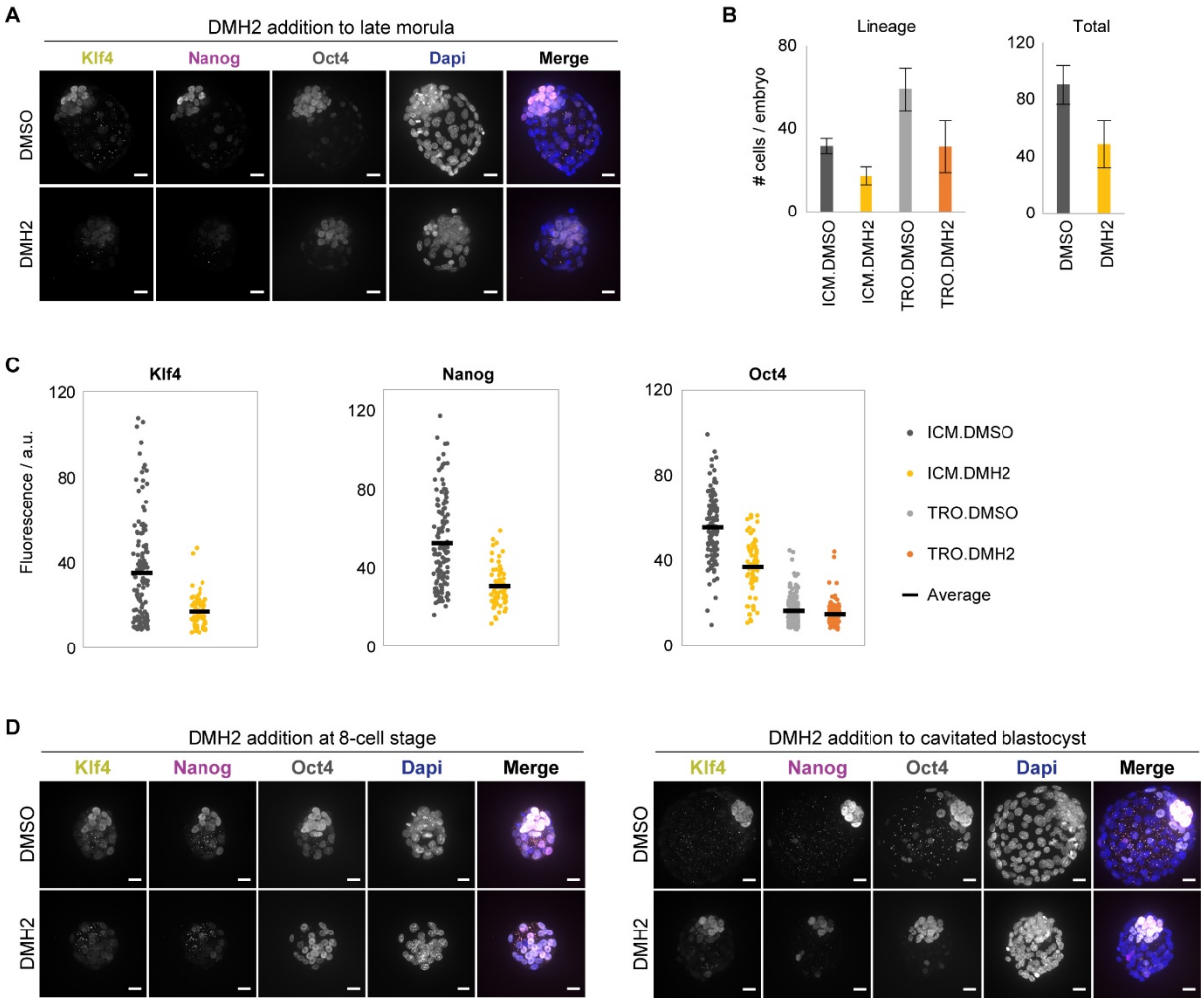


Figure 6.2.7: Timed addition of DMH2 to embryos disrupts pluripotency induction

(A–C) Mouse strain 129 compacted morulae were cultured with 3 µM DMH2 or 1:1000 DMSO from cavitation onset for 36 h. n=9.
(A) Immunofluorescent staining against Klf4, Nanog and Oct4. Representative maximum intensity projections of Z-stack 40x slices are presented. Scale bars: 20 µm.
(B) Mean cell number per embryo ±SD is shown for ICM and TRO lineages or whole embryos
(C) Quantification of immunofluorescent signal in the ICM and trophectoderm (TRO) compartments.
(D) Representative maximum intensity projections of Z-stack 40x slices for embryos to which 3 µM DMH2 or 1:1000 DMSO was added at different timepoints. Scale bars: 20 µm.
The immunofluorescent quantification was conducted in collaboration with Carla Mulas.

6.3 Discussion

6.3.1 Mechanistic distinctions between reprogramming routes

The results in Chapter 6 provide evidence that iKlf2, iPStat3 and iEsrrb drive reprogramming by mechanistically distinct routes in terms of their genetic, exogenous signal and paracrine signal requirements (Figure 6.3.1). Together with the differences in their transcriptional trajectories (Chapter 5), this demonstrated that there are multiple distinct routes by which to transition from EpiSC to iPSC identity. Therefore, this work supports theories of cell identities as multidimensional attractors (Kauffman, 1993), and reveals considerable mechanistic flexibility for the specification of a single identity from a single origin. To the best of my knowledge at the time of writing, this is the first time that mechanistic attributes have been included in consideration of whether cell identity attractors are multidimensional, rather than focussing solely on transcriptional differences.

Functional requirements:

Driver	BMP required?	LIF dependency	Chiron & PD03	Minimal	Klf2 required?
iPStat3	Yes	N/A	Synergistic	–	Yes
iEsrrb	No	High; synergy	Synergistic	LIF	No
iKlf2	Yes	Low	PD redundant	Chiron	N/A

Figure 6.3.1: Summary of mechanistic requirements for each reprogramming route

In addition to distinct sets of exogenous signal requirements, each EpiSC reprogramming driver also differentially modulated signal interpretation. Examples of this complex relationship included: transient induction of and dependence on BMP signalling for iPStat3 and iKlf2 but not iEsrrb; synergy between PD03 and Chiron in the context of PStat3 induction but PD03 redundancy when Klf2 was driving; amplification of LIF/Stat3 input to the naïve network by Esrrb (Figures 6.2.2–6.2.4). iPStat3 sensitivity to Klf2 KD was partially alleviated in the absence of PD03 (Figure 6.2.1E), reminiscent of the ability to rescue embryo-derived Nanog-/- EpiSC reprogramming in Chiron+LIF but not 2iLIF (Figure 3.2.10). Notably, both Nanog and Klf2 are considered targets of PD03 (Silva et al., 2009; Yeo et al., 2014): it appears they are required to transduce its positive input to the naïve network and, in their absence, PD03 is actively detrimental to reprogramming. Hence, interplay between TFs and signals can fundamentally modulate each other's role in identity specification.

Together, Chapters 3 and 6 exposed diverse mechanisms for amplification of exogenous LIF signal transduction by nuclear factors. Whilst Nanog suppressed Socs3 transcription and increased the amount of PStat3 (Figures 3.2.1–3.2.2), Esrrb cooperated with minute amounts of PStat3 to synergistically upregulate PStat3 targets (Figure 6.2.3). Whilst Esrrb is considered a direct target of Nanog (Festuccia et al., 2012), this point highlights that they cannot be considered as functionally redundant, and may shed light on their disparate reprogramming efficiencies (Figure 4.2.8C).

The morphological changes following iKlf2, iPStat3 or iEsrrb induction were also driver-specific (Figure 5.2.1C). This provides further evidence for distinct reprogramming routes. In the future, it would be interesting to re-interrogate the RNAseq data for clues as to how these different morphologies arise, for example by considering genes involved in regulation of the cytoskeleton and the extracellular matrix. These physical differences may be more than simple reflections of distinct intermediate transcriptomes, but could also play an active role in the modulation and interpretation of gene expression and signalling mechanisms.

6.3.2 BMP signalling in naïve pluripotency establishment

I found transient BMP signalling activation and requirement during reprogramming driven by iKlf2 and iPStat3, but not iEsrrb. This provided further evidence of route-specific mechanistic attributes, and formed an interesting parallel with recent work by Onishi et al., 2014 in which they showed that addition of BMP4 increased EpiSC reprogramming efficiency. They attributed the positive effect of BMP4 to facilitation of LIF signal transduction through GP130/LIFR and to formation of a Stat3/Smad1 protein complex together with the P300 transcriptional coactivator. In light of the latter point, I wonder whether Esrrb-driven reprogramming did not require BMP signalling due to its own ability to cooperate with PStat3 at its downstream genetic targets (Figure 6.2.3). It would be interesting to investigate whether Esrrb and PStat3 also form an activatory protein complex. Indeed, a very recent publication found that Esrrb acts as a pioneering TF during EpiSC reprogramming, binding to closed chromatin and recruiting P300 in a LIF/Stat3-dependent manner (Adachi et al., 2018). Local chromatin remodelling ensues, rendering it accessible for subsequent binding of Oct4/Sox2/Nanog at naïve loci. This mechanism is consistent with the findings I report here.

By application of insight derived from murine EpiSC reprogramming, I found that BMP signal inhibition abolished resetting of primed human ESCs towards naïve pluripotency. To clarify whether this requirement was transition-specific in human as in the murine context, it would

be important to test the BMP signalling status in established naïve human iPSCs. Some signalling differences have already been reported between naïve human and mouse pluripotent cells (Guo et al., 2017; Takashima et al., 2014; Theunissen et al., 2014). Whilst further determination of the signals governing naïve human pluripotency maintenance was beyond the scope of this thesis, I am curious to see how our understanding of this is resolved by ongoing work in different laboratories. Similarly, I hope that the findings of Figure 6.2.7 prompt reinvestigation into BMP signalling roles and mechanisms during pre-implantation embryonic development.

CHAPTER 7:
ROUTE RECONCILIATION WITH A UNIFYING TRANSITION LOGIC

7.1 Introduction

The aforementioned differences in transcriptional trajectories, signal and genetic requirements demonstrated that iKlf2, iPStat3 and iEsrrb instruct reprogramming by distinct mechanisms (Chapters 5–6). Given that the starting and destination cellular identities were the same in all three cases, the extent of the route differences was surprising. Therefore, in Chapter 7 I ask whether there is a common feature, which could reconcile the disparate transition logics.

7.2 Results

7.2.1 Defined Oct4 level is a common feature of all routes

In light of earlier observations that Oct4 drop was rescued in the *Rex1::*dGFP+ population of iK2N-driven reprogramming (Figure 4.2.3F), I checked the behaviour of Oct4 expression during transitions driven by iEsrrb, iPStat3 and iKlf2. I found that Oct4 was maintained at endogenous level throughout productive reprogramming, irrespective of the driver. In the scRNAseq of productive intermediates for each line, cells expressed Oct4 within a narrow range around EpiSC/iPSC levels (Figure 7.2.1).

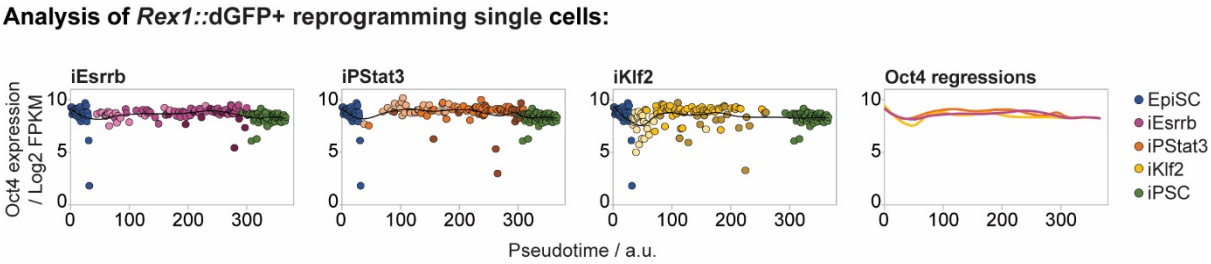


Figure 7.2.1: Oct4 maintenance is a unifying feature of productive reprogramming

Scatter plots of Oct4 expression in single cells vs pseudotime (Fig. S3C), fitted with LOESS regression lines. Regression lines are overlaid on the right for comparison. The RNAseq analyses were conducted in collaboration with Giuliano Stirparo.

Maintenance of Oct4 throughout the transitions is not to be taken for granted. Although Oct4 is expressed at similar levels in EpiSCs and iPSCs (Figure 7.2.2A), this expression is supported by different transcriptional networks and is driven from different enhancer elements (Tesar et al., 2007; Yeom et al., 1996). Indeed, signal switch of control EpiSCs to 2iLIF triggered Oct4 downregulation (Figure 7.2.2B). In contrast, Oct4 was unperturbed in ESCs upon switch from serum+LIF to 2iLIF (Figure 7.2.2C), indicating that 2i itself did not suppress Oct4 in a context where cellular identity was constant.

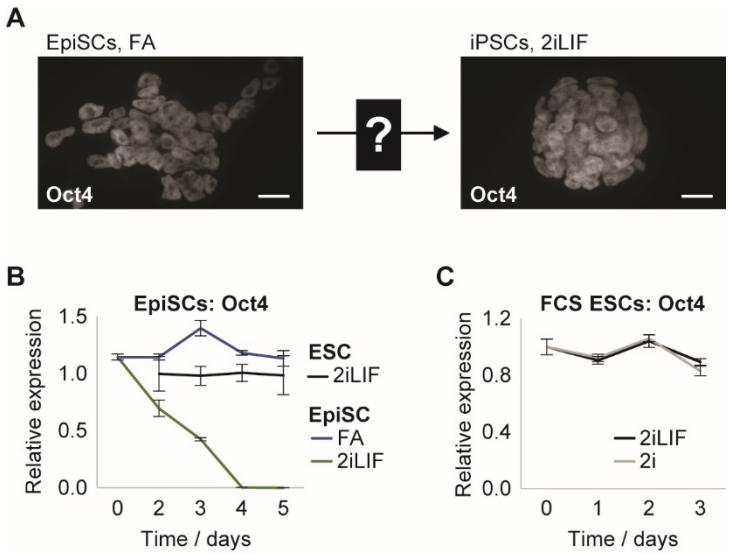


Figure 7.2.2: Oct4 maintenance during reprogramming is not to be taken for granted

(A) Immunofluorescent staining of Oct4 expression in *Rex1::*dGFP EpiSCs and iPSCs. Representative maximum intensity projections of Z-stack 40x slices are presented. Scale bars: 20 μm.
(B) Timecourse RT-qPCR analysis of EVrtTA3 control EpiSCs in FA or 2iLIF+dox, and of ESCs maintained in 2iLIF. Mean Oct4 expression is displayed relative to Gapdh and normalised to ESC day 2, ±SD (n=3).
(C) Timecourse RT-qPCR of ESCs previously cultured in FCS+LIF then switched to 2i±LIF. Mean Oct4 expression is displayed relative to Gapdh and normalised to day 0, ±SD (n=3).

7.2.2 Oct4 maintenance is an active feature of productive reprogramming

Timecourse RT-qPCR analysis showed that Oct4 was always expressed at or above ESC level in the dGFP+ reprogramming subpopulation but not always in the dGFP- (Figure 7.2.3A). Together, this suggested that signal-mediated collapse of the primed network prior to naïve network construction led to Oct4 expression loss, creating a 'vulnerable window' between different self-renewing Oct4-supporting configurations. Since 2iLIF triggered Oct4 collapse in control EpiSCs, I reason that the observed maintenance of Oct4 in 2iLIF during productive reprogramming is an active process coordinated by the driving transgene (Figures 7.2.1–7.2.2).

To evaluate the relationship between productive identity transition and Oct4 maintenance, I subdivided intermediate populations based on a finer gradient of dGFP level. Average Oct4 expression positively correlated with reprogramming efficiency of the given subpopulation (Figure 7.2.3B). To test whether this Oct4 maintenance was required for reprogramming, I performed transient Oct4 KD by a single pulse of siRNA treatment at reprogramming onset. iPSC formation was completely abolished (Figure 7.2.3C).

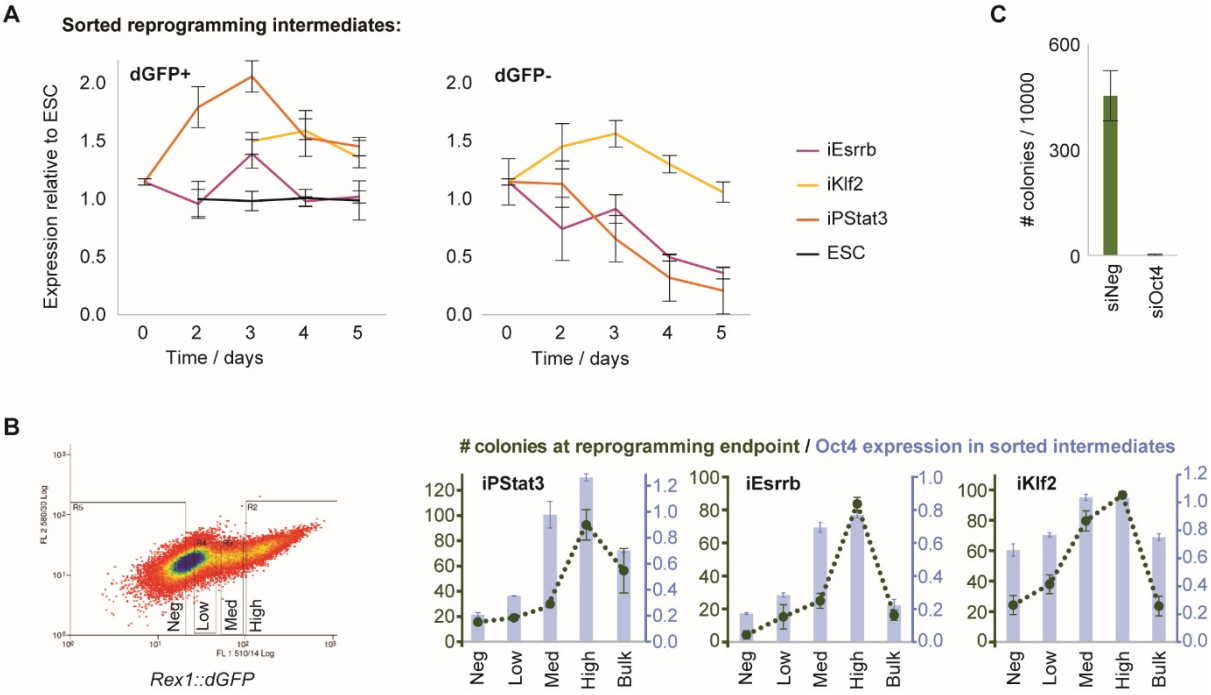


Figure 7.2.3: Oct4 maintenance correlates with and is required for reprogramming

(A) Timecourse RT-qPCR analysis of *Rex1::*dGFP+ and dGFP- EpiSC reprogramming intermediate populations in 2iLIF+dox/GCSF, and *Rex1::*dGFP+ ESCs in 2iLIF. Mean Oct4 expression is displayed relative to Gapdh and normalised to ESC day 2, ±SD (n=3). It is of interest to note that iKlf2 dGFP- cells also maintain Oct4 expression, and retain the potential to productively form iPSC colonies until day 4, unlike other dGFP- populations (data not shown).
(B) dGFP negative, low, medium, high and bulk reprogramming intermediate populations were isolated by flow cytometry, analysed for mean Oct4 expression level by RT-qPCR (blue), then replated for clonogenicity assay in 2iLIF (green). The FACS-plot on the left shows dGFP levels during reprogramming, taking the example of iEsrrb 2iLIF+dox day 3. dGFP high, medium (med), low, and negative (neg) gates are indicated. dGFP negative and high gates were set according to *Rex1+/dGFP* EpiSCs and ESCs respectively. Low and medium gates subdivide the intervening levels. This was conducted in collaboration with Tim Lohoff.
(C) Oct4 KD was performed at iPStat3 reprogramming onset with a single 16 h pulse of siRNA. iPSC colonies were scored on day 8, presented as mean ±SD (n=3).

7.2.3 Fixed Oct4 expression is sufficient for naïve instruction in minimal conditions

Having demonstrated that Oct4 maintenance was observed in and required for productive reprogramming, next I asked whether Oct4 maintenance was sufficient. *Oct4* null EpiSCs were generated that constitutively expressed ectopic Oct4 at endogenous level (FixedOct4) (Figure 7.2.4A–D), according to methodology described in Radzisheuskaya et al., 2013. This uncoupled Oct4 expression from identity or environmental perturbations. An *Oct4* null background was necessary to ensure maintenance of total Oct4 level, and to avoid overexpression of Oct4 that triggers differentiation (Niwa et al., 2000). Correspondingly, ectopic Oct4 expression on top of a wild-type background gave very inefficient EpiSC reprogramming (data not shown).

Following medium switch to 2iLIF, FixedOct4 EpiSCs rapidly generated naïve iPSC colonies with extremely high efficiency (Figure 7.2.4B–D). FixedOct4 +LIF was the minimal requirement for naïve pluripotency specification (Figure 7.2.4E). In FixedOct4 reprogramming, impetus towards the naïve identity was provided only by exogenous signals: Oct4 is expressed equally in both EpiSCs and iPSCs, so there was no naïve-specific transgene. Therefore, maintenance of Oct4 permitted the identity transition, while signals specified the direction.

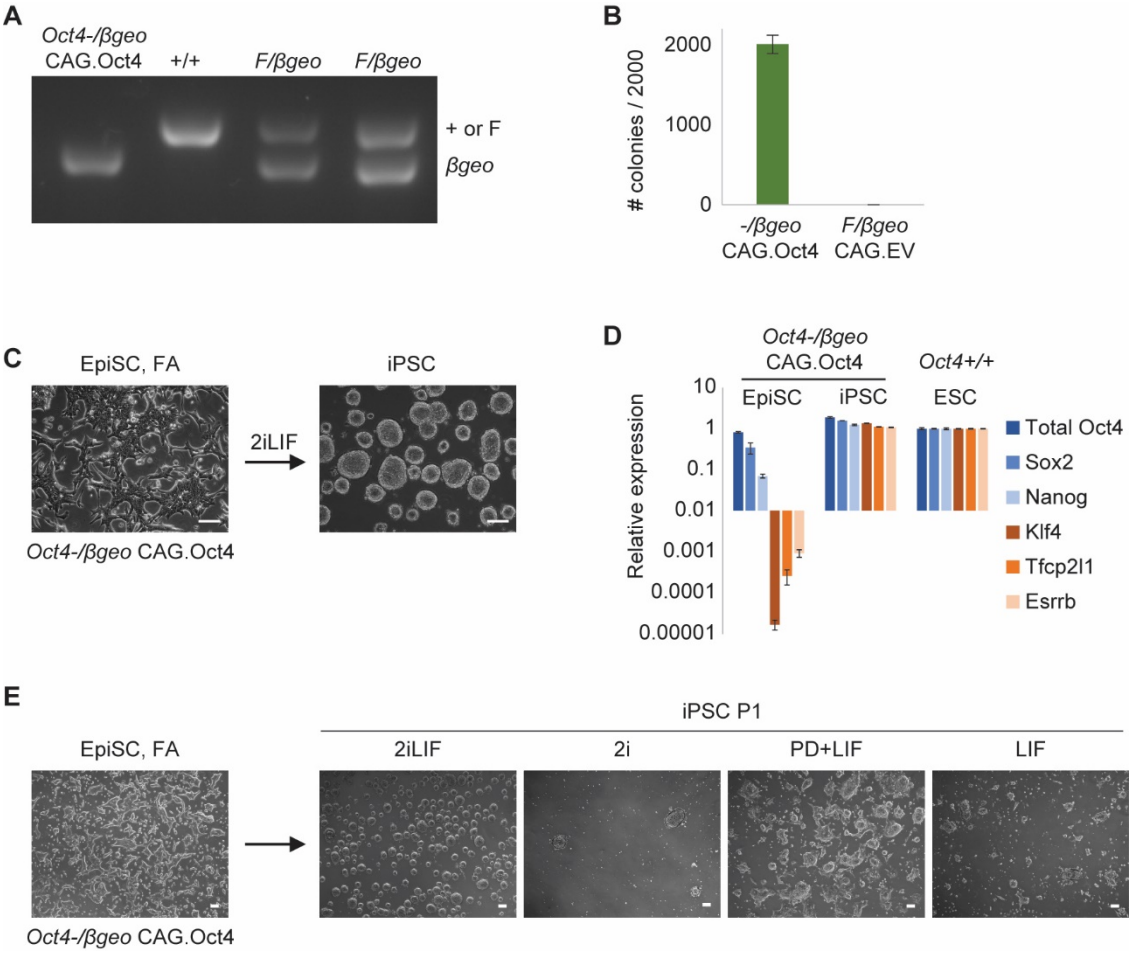


Figure 7.2.4: Fixed Oct4 expression is sufficient for naïve instruction

(A) Genotyping of *Oct4-/βgeo* CAG.Oct4 EpiSCs, with wild-type (+/+) and *Flox (F)/βgeo* controls.
(B–D) *Oct4-/βgeo* CAG.Oct4 (FixedOct4) EpiSCs formed naïve iPSC colonies at high efficiency in 2iLIF, quantified by mean iPSC colonies scored on day 8 ±SD (n=3) **(B),** indicated morphologically **(C)**, and confirmed by RT-qPCR expression analysis relative to Gapdh then normalised to ESC level=1 ±SD (n=3) **(D)**. Scale bars: 100 µm.
(E) *Oct4-/βgeo* CAG.Oct4 EpiSCs were plated at 2000/24well in FA. The following day, medium was changed to N2B27 ± Chiron ± PD03 (PD) ± LIF in all permutations (n=3). After 4 days, G418 was applied to select for endogenous *Oct4* promoter activity. Quantification of reprogramming efficiency was inappropriate in this case, since some conditions are permissive compared to 2iLIF. Instead, on day 8, each 24well was passaged in its entirety to 2 x 6wells, one maintaining the condition +G418 and one swapped to 2iLIF+G418 to challenge iPSC clonogenicity in this naïve-selective condition. iPSC generation propensity was assessed on day 12, and representative condition +G418 phase images are shown for those conditions which successfully generated iPSCs. Scale bars: 100 µm. We found that FixedOct4 +LIF was the minimal requirement for naïve pluripotency specification, with derivative iPSCs expandable in LIF or 2iLIF+G418 for at least 8 passages. Curiously, Chiron+LIF did not instruct naïve pluripotency acquisition for FixedOct4 EpiSCs despite initial emergence of naïve-like morphology (n=10).

7.2.4 Reconciliation of route differences with common Oct4 maintenance

Despite the distinctions between routes in terms of their transcriptional trajectories and mechanistic requirements (Figures 5.3.1, 6.3.1), Oct4 maintenance was a common feature and I further demonstrated that this is required and sufficient for reprogramming (Figure 7.2.1–7.2.4). Now I integrate route-specific attributes with the common denominator of Oct4 maintenance, to address the overarching logic of the transitions.

First, I assessed the ability of each factor individually to rescue the drop in Oct4 expression when EpiSCs are treated with 2iLIF for 24 h (Figure 7.2.5A). Klf2 induction yielded the most effective rescue of Oct4 drop, including on the protein level (Figure 7.2.5A–D). This Oct4 support may explain the high efficiency of Klf2-driven reprogramming despite its paradoxical dearth of naïve gene induction (Figure 5.2.1A). iPStat3 somewhat maintained Oct4 expression. However, the remaining drivers failed to rescue the Oct4 drop in bulk population (Figure 7.2.5A).

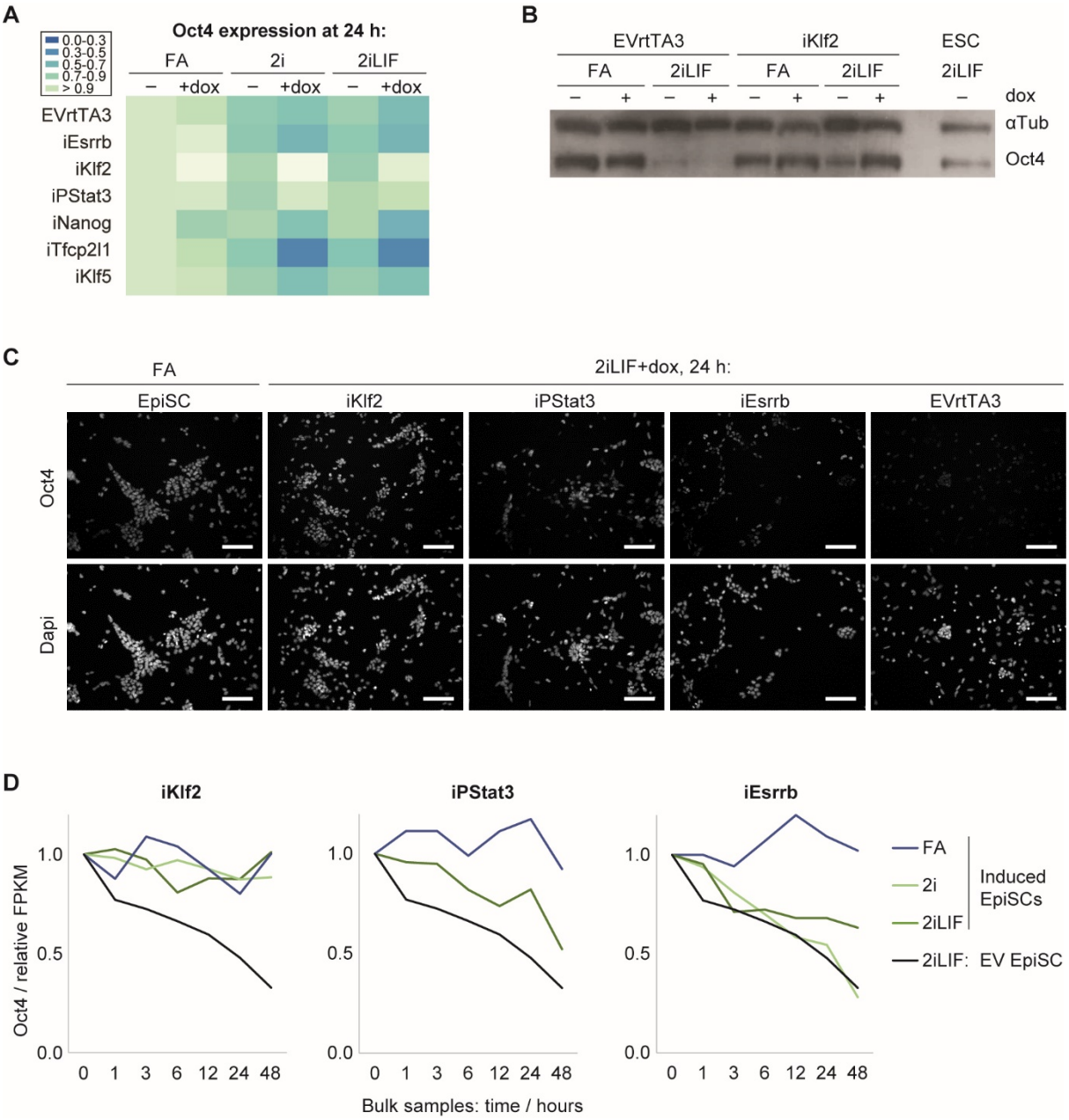


Figure 7.2.5: Immediate impact of each driver on Oct4 expression

(A) Heatmap of Oct4 expression after 24 h of 2iLIF+dox/GCSF treatment, measured by RT-qPCR relative to Gapdh then normalised to EpiSC-level=1.
(B) Western blot against Oct4 and αTubulin in EVrtTA3 and iKlf2 EpiSCs, after 24 h in FA or 2iLIF ± dox.
(C) Immunofluorescent staining of Oct4 expression in *Rex1::*dGFP iKlf2, iPStat3, iEsrrb and EVrtTA3 EpiSCs, after 24 h in 2iLIF+dox/GCSF. Representative 20x images are presented. Scale bars: 100 µm.
(D) Timecourse gene expression analysis of iKlf2, iPStat3, iEsrrb and EVrtTA3 EpiSCs induced in the indicated conditions +dox/GCSF. Oct4 expression is presented relative to 0 h. This was conducted in collaboration with Tim Lohoff.

Since Klf2 was the most effective supporter of Oct4 (Figure 7.2.5), an early transcriptional responder to iPStat3 (Figure 6.2.1A), and since transient Klf2 KD abolished iPStat3-driven reprogramming (Figure 6.2.1B–D), I asked whether these observations could be conceptually integrated. Transient Klf2 KD at iPStat3 reprogramming onset resulted in a 50% reduction of Oct4 expression (Figure 7.2.6A). In contrast, Klf2 KD did not abolish reprogramming of FixedOct4 EpiSCs (Figure 7.2.6B), even though this was a highly LIF/Stat3-dependent process (Figure 7.2.4E). This places Oct4 maintenance as a functionally important downstream mechanism of Klf2 in reprogramming, likely to be direct due to its manifestation within one hour (Figure 7.2.5D).

iEsrrb was the most efficient of all tested drivers (Figure 4.2.8C), yet exhibited an initial drop in Oct4 expression at 24 h (Figure 7.2.5) prior to recovery in the productive subpopulation by 48 h (Figures 7.2.1, 7.2.3A). The outstanding feature of iEsrrb reprogramming initiation kinetics was the rapid and strong upregulation of naïve network genes in a highly 2iLIF-dependent manner (Figures 5.2.1A, 6.2.3). To test whether this corresponded to rapid wiring of a coherent self-renewing naïve network, I challenged the transgene-independent clonogenicity of iEsrrb *Rex1*::dGFP+ cells at 48 h by replating single sorted cells in 2iLIF without dox. Strikingly, their dox-independent clonogenicity was comparable to ESCs (Figure 7.2.6C), indicating that 48 h post-induction a functional naïve network has already formed for iEsrrb. Thus I propose that iEsrrb drives a rapid transition between primed and naïve networks, reducing the duration of the 'vulnerable window' for Oct4 expression loss between different self-renewing states.

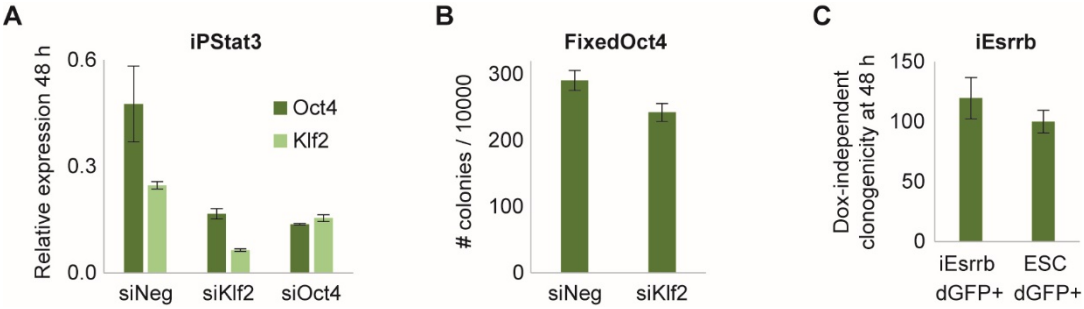


Figure 7.2.6: Reconciliation of naïve induction logic between different routes

(A) iPStat3 EpiSCs were treated with a single pulse of siRNA at reprogramming onset in 2iLIF+GCSF. After 48 h, gene expression was analysed by RT-qPCR relative to Gapdh then normalised to iPSC-level=1.
(B) Klf2 KD was performed at reprogramming onset of FixedOct4 EpiSCs in 2iLIF. G418 was applied at day 4, and iPSC colonies scored on day 8, presented as mean ±SD (n=3).
(C) dGFP+ ESCs and iEsrrb reprogramming intermediates were isolated by FACS on day 2 of reprogramming in 2iLIF+dox and plated for clonal assay in 2iLIF without dox. Blasticidin was applied on day 6 and naïve colonies scored on day 9. Mean colony number is presented as a percentage of ESC colonies, ±SD (n=3).

I tested whether this Oct4-supportive role of Klf2 was conserved in human cells. As in murine EpiSCs, hOct4 expression dropped dramatically upon switch from primed medium to 2iLIF+dox (Figure 7.2.7A). hOct4 drop was partially rescued by iKlf2, whereas hSox2 was suppressed (Figure 7.2.7A–B), suggesting a similar role of Klf2 in human and mouse reprogramming (Figures 5.2.1A, 7.2.5D).

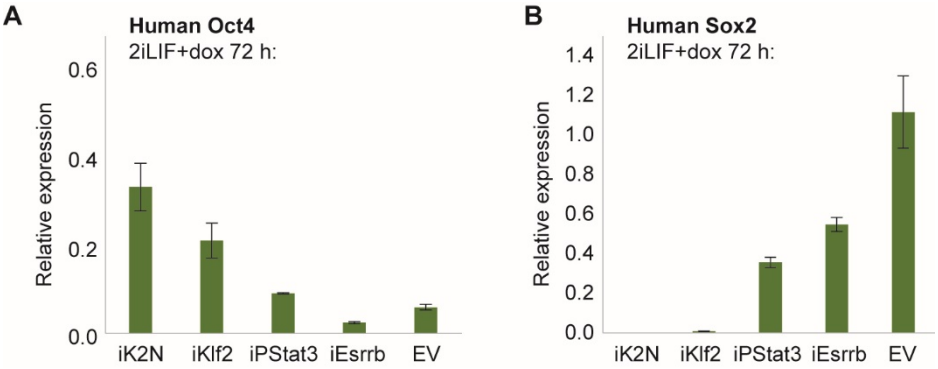


Figure 7.2.7: Conservation of Oct4 vs Sox2 regulation between species

RT-qPCR analysis of human H9 ESCs after 72 h of transgene induction in 2iLIF. Mean Oct4 **(A)** or Sox2 **(B)** expression ±SD (n=3) is displayed relative to Gapdh, normalised to EV 0 h. This was conducted in collaboration with Tim Lohoff.

7.2.5 Fine-tuned Oct4 expression is a pivotal feature for somatic reprogramming

To address the applicability of these findings to other cellular contexts, I derived somatic cells from FixedOct4 iPSCs by differentiation in chimeras (Figure 7.2.8A). Contribution of FixedOct4 cells to chimeras was high, and their Oct4 expression persisted (Figure 7.2.8B). I confirmed bonafide developmental progression by loss of Nanog, expression of lineage markers in the appropriate locations, and histological analyses (Figure 7.2.8B–C, Appendix page 187). This is not so surprising as it may seem: there is precedent for Oct4-expressing cells to proceed in development. (Masui et al., 2007; Radzisheuskaya et al., 2013; Ramos-Mejía et al., 2005).

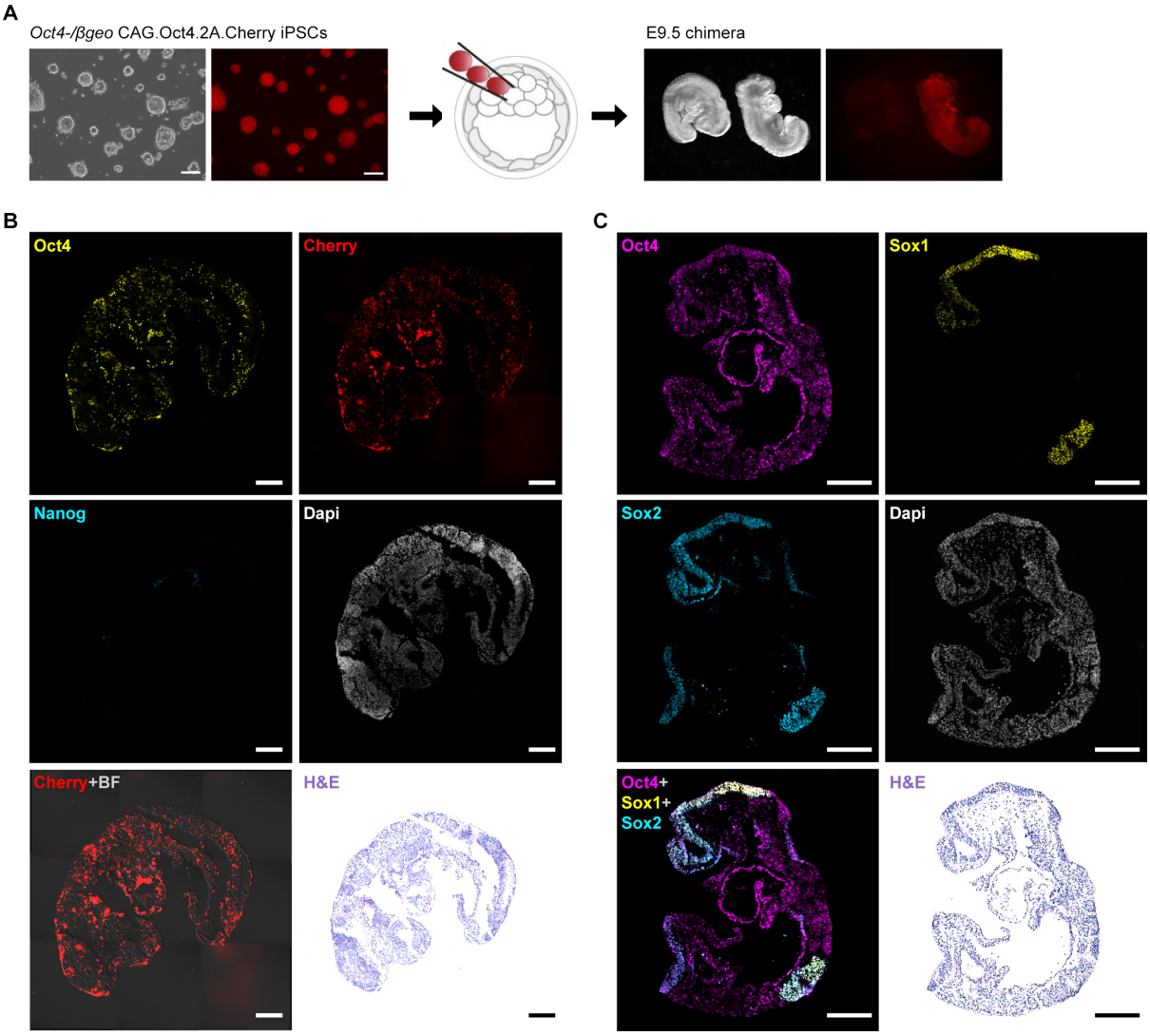


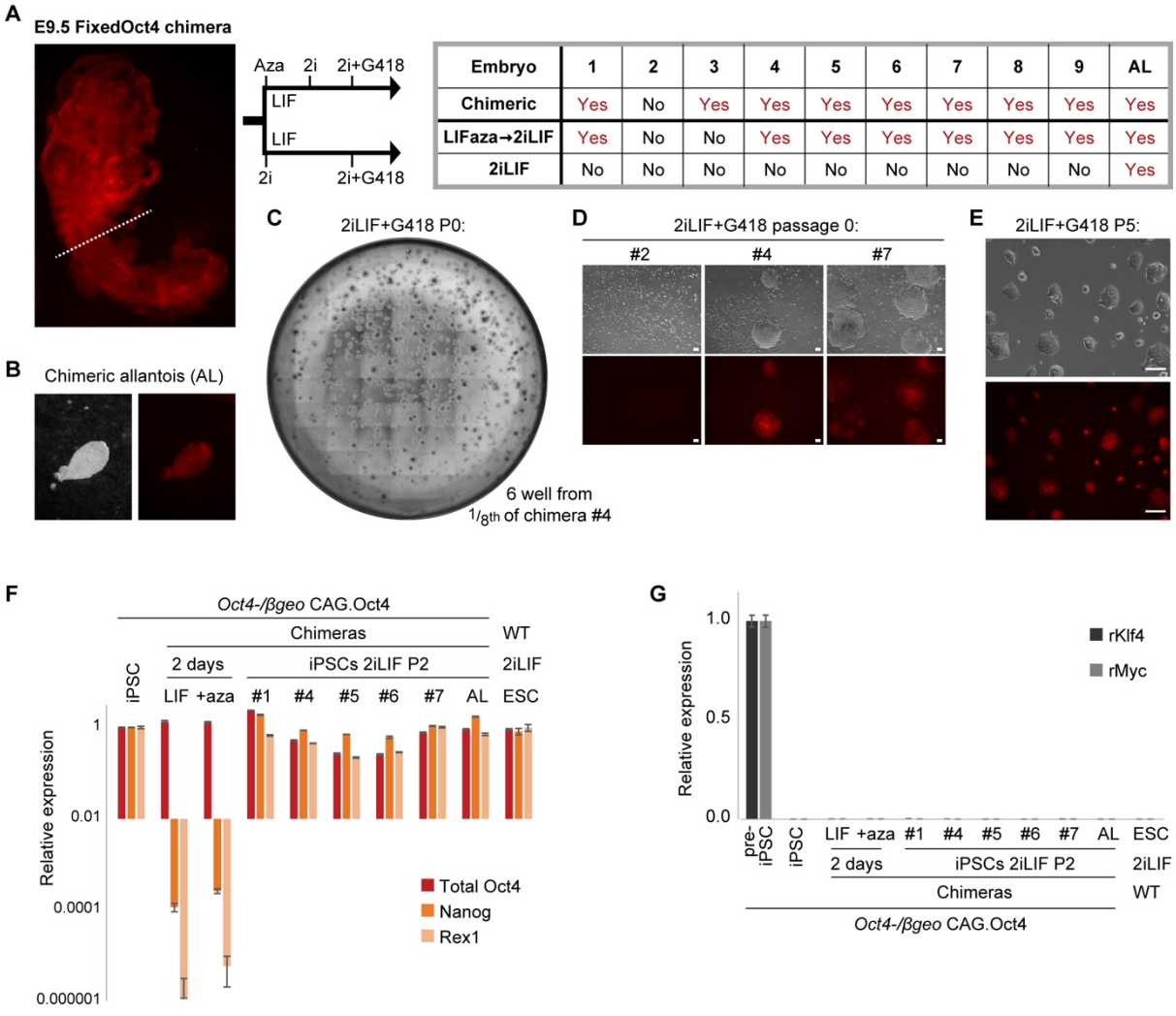
Figure 7.2.8: Derivation of FixedOct4 somatic cells by differentiation in chimeras

(A) *Oct4-/βgeo* CAG.Oct4.2A.Cherry iPSCs were injected into E3.5 C57BL/6 blastocysts then transferred to recipients. Resultant embryos were collected at E9.5. Representative phase and Cherry images are shown. Scale bars: 100 µm.
(B–C) Marker expressions were analysed by immunofluorescence on 8 µm cryosections. Subsequently, histological analyses were performed by H&E staining. Scale bars: 200 µm.

I discarded a generous portion of the tail to stringently avoid germ cell contamination, then dissociated and cultured the anterior portion of each chimera to test reprogramming ability (Figure 7.2.9A). In general, reprogramming from somatic cells takes considerably longer than from EpiSCs, which can be explained by the need to remodel a more constrained epigenetic landscape. For this reason, conversion of FixedOct4 somatic cells was attempted using two different strategies: directly in 2iLIF, or in LIF combined with a low dose of 5-Azacytidine (aza, an inhibitor of DNA methyltransferase activity) prior to 2iLIF treatment.

With the exception of positive control allantois (AL), iPSCs were not generated when plated directly in 2iLIF, consistent with previous demonstrations that, if applied from the beginning, 2iLIF does not support somatic cell reprogramming (Silva et al., 2008). Strikingly however, naïve iPSCs were generated at high efficiency following culture in LIFaza then 2iLIF (Figure 7.2.9A–G). This indicates that, as in EpiSC reprogramming (Figure 7.2.4), fine-tuned Oct4 expression was able to drive efficient naïve pluripotency establishment from somatic cells[8].

[8] In subsequent experiments, I challenged whether aza is required to overcome epigenetic barriers from more advanced developmental states. Aza is not essential. FixedOct4 together with Jak/Stat signalling is sufficient for naïve pluripotency induction from E9.5 cells and E12.5 MEFs (Appendix page 187).



Embryo	1	2	3	4	5	6	7	8	9	AL
Chimeric	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LIFaza→2iLIF	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2iLIF	No	No	No	No	No	No	No	No	No	Yes

Figure 7.2.9: Fine-tuned Oct4 expression is a pivotal feature for somatic reprogramming

(A) The tail portion of E9.5 embryos was removed to strictly avoid germ cell contamination. The anterior portion was dissociated manually, then subdivided into quarters (2x LIFaza, 2x 2iLIF). After 6 days, LIFaza was exchanged for 2iLIF, then on day 10 G418 was applied to all cultures. A chimeric allantois (AL) (Fig. S7B) was plated as a positive control (germ-cell containing) directly in 2iLIF. Generation of G418-resistant iPSCs is summarised in the table.
(B) Phase and Cherry images of allantois (AL) dissected from a headfold stage embryo.
(C) Representative phase image showing G418-resistant iPSCs in an entire 6well at passage 0, following LIFaza then 2iLIF treatment on 1/8th of a chimeric embryo.
(D) Representative phase and Cherry images at passage 0 in 2iLIF+G418, following derivation in LIFaza then 2iLIF. Scale bars: 100 µm. This provides a closer view of the timepoint in **(C)**.
(E) Representative phase and Cherry images of iPSCs at passage 5 in 2iLIF+G418, following derivation from chimera #4 in LIFaza then 2iLIF. Scale bars: 100 µm.
(F–G) RT-qPCR expression analysis relative to Gapdh then normalised to parental iPSC level=1 **(F)** or to pre-iPSC level=1 **(G)**. Expression is shown after two days of treatment with LIF±aza on cells from chimera #1. Expression of iPSCs at passage 2 in 2iLIF+G418 is shown after derivation in LIFaza then 2iLIF for chimeras #1, 4–7, and after derivation directly in 2iLIF for the allantois (AL). Due to the iPSC-history of the cells (Radzisheuskaya et al., 2013), it was necessary to confirm that silenced retroviruses (r) were not reactivated.
This was conducted in collaboration with Lawrence Bates.

7.3 Discussion

7.3.1 Unification of reprogramming logics

During reprogramming initiation, Oct4 expression dropped as naïve signals disrupted the primed network (Figures 4.2.1, 4.2.2, 7.2.2, 7.2.5). However, Oct4 expression was maintained throughout reprogramming in the productive subpopulation, irrespective of the driving factor (Figures 4.2.3, 7.2.1, 7.2.3), whereas Oct4 knock-down abolished reprogramming (Figure 7.2.3). Therefore, Oct4 drop following signal switch presented a barrier that had to be actively overcome by the drivers for productive reprogramming. The functional importance of this Oct4 maintenance was validated by fixing Oct4 to endogenous level (Figure 7.2.4), which was sufficient for extremely efficient entry to the naïve identity under only signal instruction.

Therefore, distinct routes of reprogramming (Figures 5.3.1, 6.3.1) can be thought of as different strategies by which to achieve the unifying, required and sufficient event of Oct4 maintenance (Figure 7.3.1). In light of this I propose the following hypothesis: for a given EpiSC reprogramming driver, there is a certain probability of rescuing Oct4 expression during the 'vulnerable window' after Oct4 loses support from the collapsing primed network. I suggest that reprogramming efficiency correlates with this probability, which is determined by the ability of that factor itself to drive Oct4 expression, and the speed at which that factor orchestrates a coherent naïve network to support Oct4 in an alternative topology. iKlf2 and iEsrrb occupy opposite extremes within this model, relying solely on the former and latter methods respectively (Figures 7.2.5, 7.2.6). iPStat3 employs an intermediate strategy, eliciting moderate upregulation of naïve network genes and with Klf2 acting downstream of PStat3 to support Oct4 expression (Figure 7.2.5). This lends further significance to my identification of Klf2 as a PStat3 target in EpiSCs despite not being considered as such in ESCs (Figures 3.2.9, 6.2.1): not only is network topology different during naïve pluripotency induction vs maintenance, but this has important functional consequences.

The permissible time-period and level of transient Oct4 drop should be clearly delineated prior to its recovery at 48 h, particularly for iEsrrb (Figure 7.2.1, 7.2.3, 7.2.5). Since tagging of Oct4 renders it dysfunctional (Silva group, unpublished), this could instead be attempted by creation of a destabilised reporter at an endogenous *Oct4* locus, connected between the open-reading-frame and 3' UTR by an IRES or 2A sequence.

7.3.2 Applicability to somatic cell reprogramming

Results in a somatic cell context supported my findings in EpiSCs and further demonstrated that identity transition to naïve pluripotency pivots on the fine-tuning of Oct4 (Figure 7.2.9). It was previously shown that Oct4 is sufficient to induce pluripotency from somatic cells (Kim et al., 2009), but this was very inefficient and protracted (0.014%, 4–5 weeks). Instead, fixing Oct4 expression to an ESC level resulted in efficient and rapid somatic reprogramming (Figure 7.2.9). Conversely, somatic reprogramming blocked in the final stage by lack of endogenous Oct4 was rescued by provision of ectopic Oct4 at control level (Radzisheuskaya, 2014). Thus, precise Oct4 expression is a defining feature in distinct contexts of nuclear reprogramming. I suggest that correct Oct4 level creates the opportunity for transition into naïve pluripotency, which is effected provided there is a conducive environment.

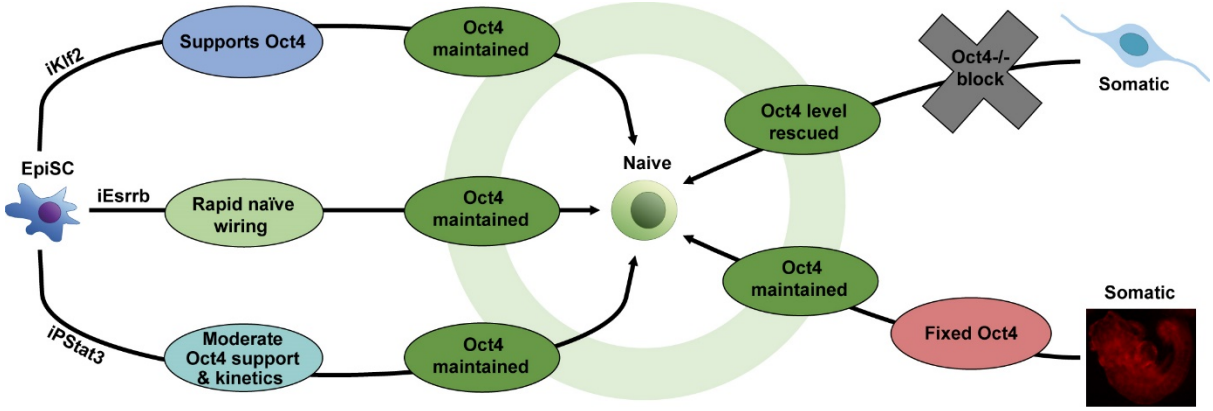


Figure 7.3.1: Convergence on precise Oct4 expression to access naïve pluripotency

Schematic summarising the convergent feature of correct Oct4 level, which permits cells to transit into naïve pluripotency. Left: diverse logics by which different EpiSC reprogramming drivers achieve the required common feature of Oct4 maintenance in order to undergo the identity transition. Right: in distinct somatic cell reprogramming contexts, expression level of Oct4 is a key and limiting mechanism in order to access the naïve identity. From both EpiSCs and E9.5 somatic cells, FixedOct4+LIF is sufficient to drive reprogramming to naïve pluripotency.

7.3.3 Resolving the iKlf2 paradox

iKlf2 is rather enigmatic as an efficient EpiSC reprogramming driver (Figure 4.2.8C). Its dearth of naïve gene upregulation within the first 48 h is counterintuitive, as is its highly divergent initiation trajectory and upregulation of mesodermal markers even in productive single cells (Figures 5.2.1, 5.2.5). In the first 48 h of reprogramming, the only positive impact of Klf2 on pluripotency genes is robust support of Oct4 expression (Figures 5.2.1, 7.2.5). Since FixedOct4 is sufficient for highly efficient reprogramming (Figure 7.2.4), I reason that a similar phenomenon happens here: iKlf2 intermediates are Oct4+ and thus remain permissive for reprogramming directed by signals. The slow kinetics with which naïve genes are eventually upregulated in productive iKlf2 cells is consistent with gradual network coordination as signal instruction becomes dominant (Chapter 5). I note that Oct4 expression is initially maintained during mesendoderm lineage entry (Downs, 2008; Thomson et al., 2011) and thus speculate that transient diversion can be beneficial for reprogramming if it helps to achieve the Oct4 maintenance requirement. Conversely, Sox2 is downregulated during mesendoderm specification (Thomson et al., 2011). Here I report concomitant Oct4 support and Sox2 silencing in response to iKlf2 or iK2N induction, and this is conserved between murine and human systems (Figures 4.2.3, 5.2.1, 7.2.5, 7.2.7). In both species, it will now be interesting to understand how and why Oct4 expression is maintained within the transition-permissive range, and whether Sox2 suppression is of functional significance.

CHAPTER 8: GENERAL DISCUSSION

8.1 General conclusions and implications

8.1.1 Principles of cell identity transitions

In the preceding Chapters, I showed that there are multiple routes by which the naïve identity can be established from EpiSCs, with the unifying feature of active Oct4 maintenance. Not only do these routes differ in their transcriptional trajectories (Chapter 5), but crucially also in their mechanistic attributes as evidenced by exogenous signal, paracrine signal and genetic requirements (Chapter 6). Thus, I demonstrate that there can be considerable flexibility for the specification of a single identity from a single origin. This adds further complexity to the paradigm of multicellular biology in which TFs and signals are used in different permutations and contexts in order to generate different cell types: they can also be used in different ways to generate the same cell type. I wonder whether this principle could assist organismal robustness and flexibility in the face of unpredictable events.

This thesis supports theories that cellular identities are multidimensional attractors, occupying local minima of stable network states with multiple paths of approach (Section 1.1) (Huang et al., 2005; Kauffman, 1993). Here I provide a substantial advance on previous works, reaching a single destination identity via three different routes with mechanistic as well as transcriptional distinctions. Formally, an attractor and the transitions between them are computed on the basis of transcriptome signatures and underlying GRNs (Davila-Velderrain et al., 2015). Whilst modelling of routes through identity state space according to dynamical systems mathematics was beyond the scope of the thesis, I hope such analyses will be performed on our RNAseq data. In the meantime, it is clear from the genome-wide analyses in Chapter 5 that distinct transcriptional trajectories are taken. These also differ in their signal and genetic requirements (Chapter 6), demonstrating mechanistic distinction between transition routes. So far as I am aware, this is the first time that mechanistic attributes have been considered when assessing whether cellular attractors are multidimensional. Whilst such parameters are difficult to formalise in mathematical terms, in my opinion they are essential to consider from the biological perspective, validating that transitions occur via truly different intermediate cellular states.

In Chapter 7, I then reveal the logic underpinning multidimensional access to the single attractor state: fine-tuned support of a transition factor, in this case Oct4 (Chapter 7). In line

with this concept, I propose that the aforementioned distinct routes of reprogramming represent different paths of approach towards the naïve pluripotent attractor (Figure 7.3.1). In such a paradigm, I speculate that appropriate Oct4 expression would allow unstable transitioning cells to crest energetic ridges, providing access to surrounding attractors. Having met this criterion, a conducive signal environment would nudge intermediates into the naïve valley, into which they stably settle after transgene withdrawal. Induction of naïve TFs (driver and targets) expedites this approach (Figure 7.2.6C), but is not strictly necessary (Figure 7.2.4). Indeed, in the case of iKlf2 a different program is initially induced rather than that of the naïve identity (Chapter 5). This represents a conceptual shift, exposing expression of a transition factor as more important than the transcriptional program directly induced by a driver. Thus, transition of cellular identity does not simply require activation of the destination program, but instead pivots on the mechanism that permits an identity transition to occur.

An extreme example is proffered by the FixedOct4 case, where opposing yet highly efficient identity transitions occur depending solely on the environment: induction of naïve pluripotency in the presence of LIF (Figures 7.2.4, 7.2.9), or re-entry to development and bonafide contribution to embryonic lineages *in vivo* (Figure 7.2.8). Oct4 also plays a transition-permitting role during early differentiation of several lineages (Niwa et al., 2000; Radzisheuskaya et al., 2013), and can be briefly utilised to promote direct transdifferentiation from fibroblast to neural identity (Thier et al., 2012). In this light, and considering that low-Oct4 locks ESCs in self-renewal (Radzisheuskaya et al., 2013), I am now inclined to consider Oct4 as a ‘transition’ rather than a ‘pluripotency’ factor. Ultimately, since the objective of pluripotency is to execute identity transitions, perhaps this should not be so surprising.

8.1.2 Interplay between transcription factors and environmental signals

Throughout this thesis, several forms of signal/TF interplay are exemplified:

Positive interactions regarding LIF/PStat3:

1. Suppression of Socs3 and amplification of LIF signal transduction by Nanog, leading to increased PStat3 levels (Chapter 3).
2. Cooperation between LIF/PStat3 and Nanog to upregulate Klf4 and its novel enhancer RNA, termed *K4eRNA* (Chapter 3).
3. Cooperation between Esrrb and LIF to synergistically upregulate PStat3-target genes, even when only minute amounts of PStat3 are present (Chapter 6).
4. Combined action of FixedOct4 and LIF suffices to induce naïve pluripotency from EpiSCs and somatic cells (Chapter 7).
5. Forced activation of Stat3 boosts Sox2 expression in EpiSCs, and counteracts Nanog-mediated Sox2-suppression (Chapter 3).
6. Unlike in ESCs, Klf2 responds as a PStat3 target gene in EpiSCs, and this is functionally necessary for PStat3-driven reprogramming (Chapters 3, 6 & 7).

Positive interactions regarding Chiron:

7. Chiron from day 1–4 is required for Klf2 to drive EpiSC reprogramming (Chapter 6).
8. Chiron increases the efficiency of Esrrb-driven EpiSC reprogramming (Chapter 6).

Positive interactions regarding PD03:

9. PD03 increases the efficiency of FixedOct4 EpiSC reprogramming, and is required if Chiron is also present (Chapter 7).

Positive interactions regarding 2iLIF component combinations:

10. Increase in naïve network induction kinetics in 2iLIF compared to FA for iNanog+iPStat3 (Chapter 4) and iEsrrb (Chapter 5).
11. Additive effect of PD03 and Chiron on reprogramming efficiency when driven by PStat3 or iEsrrb (Chapter 6).

Positive interactions with other signals:

12. Transient dependence on BMP signalling for iPStat3 and iKlf2 (Chapter 6).

Redundancies:

13. PD03 redundancy when iKlf2 drives reprogramming (Chapter 6).

Negative interactions:

14. Rescue of *Nanog-/-* reprogramming in the absence of PD03 (Chapter 6).
15. Partial alleviation of iPStat3 sensitivity to Klf2 KD in the absence of PD03 (Chapter 6).
16. Suppression of Oct4 and Sox2 by 2iLIF treatment of EpiSCs (Chapters 4 & 6).
17. Abolishment of FixedOct4+LIF reprogramming by the addition of Chiron (Chapter 7).

Specific considerations of these findings are given in the corresponding Chapters. In a more general sense, these findings enrich our understanding of how the naïve pluripotent identity is governed by complex TF/signal interplay. This plethora of positive, negative, cooperative and redundant interactions is augmented by the ability of TFs and signals to fundamentally modulate whether and how each other is interpreted. In sum, this thesis highlights the essential role of context when considering the biological response to a given variable.

8.1.3 Network differences for installation vs maintenance of naïve pluripotency

An emergent theme from this work is that different regulatory principles apply during induction vs self-renewal of naïve pluripotency. In particular, Klf2 behaves as a PStat3-target during EpiSC reprogramming initiation (Chapters 3 & 6), but not during ESC maintenance (Hall et al., 2009). Not only does this represent a difference in network topology, but it bears important functional consequences: Klf2 is required downstream of PStat3 during reprogramming (Chapter 7). I also note suppression of Sox2 by Klf2 or Nanog induction in EpiSCs (Chapters 3 & 5). So far as I am aware, this has not been reported in ESCs.

Whilst Esrrb substitutes the requirement for Chiron in 2i culture of ESCs (Martello et al., 2012), during EpiSC reprogramming I observe synergy of Esrrb with Chiron (Chapter 4). Chiron application to EpiSCs does not upregulate endogenous Esrrb in the way that it does for ESCs (Chapters 5 & 6). When other drivers are induced, Esrrb is not upregulated early in the reprogramming process, yet Chiron application is usually beneficial (Chapter 6). Therefore, it logically follows that Chiron plays different roles during acquisition vs maintenance of naïve pluripotency. In general, Chiron signal interpretation is known to be context-dependent: as a component of 2i, Chiron supports naïve pluripotency, whereas treatment of ESCs or EpiSCs with Chiron+ActivinA elicits robust mesodermal differentiation. Since EpiSCs are already transducing ActivinA signal, it seems reasonable that their initial response to Chiron is not equivalent to that of ESCs. I hypothesize that additional, as yet unknown, inputs are required to reactivate Esrrb expression during reprogramming.

In any context, I would not tacitly assume that the rules governing identity installation and maintenance are the same, but rather challenge each process with an open mind.

8.2 Suggestions for future work

8.2.1 Genomic binding and role of Oct4 during identity transitions

An appropriate Oct4 level permits cells to transition between identities, both in differentiation and reprogramming, but the direction of the transition is instructed by other factors and signals. Primed vs naïve identities are characterised by different Oct4 binding configurations (Buecker et al., 2014). In addition to supporting Oct4 expression itself, how do other factors/signals drive reorganisation of Oct4 genomic binding? Is this a unifying feature between different reprogramming routes, or another point of mechanistic distinction? Is identity transition simply a matter of redirecting Oct4 from the start- to destination-identity binding sites, or is the role of Oct4 contingent on transition-specific events? What is the relationship between Oct4 and the epigenetic landscape? In the first instance, I would suggest ChIP studies to assess Oct4 engagement with the genome following independent manipulations of TFs and signals during reprogramming from EpiSCs to iPSCs. In particular, the ability of PStat3 to dominantly drive reprogramming in FA conditions will allow controlled comparison of 2i vs TF inputs to the redirection and action of Oct4 (Figure 4.2.1) (van Oosten et al., 2012).

Whilst the importance of Oct4 for the execution of cell identity transitions is clear, the mechanisms by which it achieves this remain elusive. Part of the difficulty in studying this lies in the very features that make the question so interesting. Oct4 performs a multitude of roles in an exquisitely dose- and context-dependent manner. In ESCs, Oct4 at zero, low, wild-type and overexpressed levels respectively trigger differentiation, locked self-renewal with inability to differentiate, context-appropriate self-renewal or differentiation, and forced differentiation (Niwa et al., 2000; Radzisheuskaya et al., 2013). Thus, even subtle manipulations can abolish the particular process that one is trying to study. For example, knock-down or conditional knock-out of Oct4 inevitably pass through an Oct4-low state, due to the time taken to degrade all the Oct4 protein, and this has utterly different functional consequences compared to either higher or lower Oct4 levels. Ongoing work (Bates et al., unpublished) may resolve this technical challenge by employing the rapid auxin-inducible degron (Nishimura et al., 2009). Similarly, a crucial benefit of our FixedOct4 system is that Oct4 is supplied at precisely the ESC-level, circumventing neomorphic effects which plague much of the existing literature (Radzisheuskaya and Silva, 2014).

8.2.2 Balancing of the naïve pluripotency network in ESCs

Robust self-renewal of ESCs must be balanced with the capacity to differentiate on demand (Chapter 1). This implies that naïve network strength must stay within certain limits. In this thesis, I have identified several negative regulatory relationships that may assist in naïve network balancing.

In EpiSCs, Sox2 is suppressed following induction of Klf2 and/or Nanog (Chapters 3–5). Conversely, PStat3 boosts Sox2, and together the inputs of PStat3+Nanog balance (Chapter 3). This may reveal a mechanism to temper the naïve pluripotent network. However, these relationships should now be carefully checked in ESCs, since responses in the EpiSC context may instead reflect onwards developmental logic. From the mid–late-streak stages, Sox2 and Nanog become anteriorly and posteriorly restricted respectively: the observed inverse relationship between Sox2 and Nanog may thus echo natural symmetry-breaking of the post-implantation epiblast.

From Chapter 4, it is apparent that ESCs have the means to limit PStat3 levels, even when total Stat3 is grossly overexpressed. In agreement with the work of Tai et al., 2014, PStat3 levels above this threshold are incompatible with long-term self-renewal of ESCs. Besides auto-induction of the Socs3 repressor, very little is known about the mechanisms of PStat3 restraint in ESCs. Since Socs-insensitive GY118F-stimulated ESCs can eventually curb PStat3 (Figure 4.2.4C), I think it would be worth reopening investigation into other repressors of Jak/Stat signalling, such as the Pias and Shp protein families (Böhmer and Friedrich, 2014).

As outlined above, a precise level of Oct4 expression is compatible with the choice between ESC self-renewal or differentiation. Levels above or below this impair one process or the other. How do ESCs constrain Oct4 within this narrow range? A variety of Oct4-supportive factors have been reported in the naïve network (Martello and Smith, 2014), yet logically they cannot form an endless positive feedback loop. What are the suppressors of Oct4, that ensure its fine-tuning? I also find it intriguing that Oct4 is expressed at the same level in FCS+LIF and 2iLIF ESC cultures, despite weaker naïve gene expression in the former (Section 1.2). In this thesis, I identified Klf2 as a potent supporter of Oct4 expression (Chapter 7). Conversely, I observed Oct4 suppression following Tfcp2l1 induction (Figure 7.2.5A), and negative correlation between Tfcp2l1 and Oct4 responses (Figure 4.2.1–2). Tfcp2l1 and Klf2 are co-induced by 2iLIF relative to FCS+LIF (Ye et al., 2013; Yeo et al., 2014), are simultaneously downregulated at implantation and reactivated during diapause *in vivo*, throughout all of which

Oct4 is maintained. Therefore, I wonder whether Tfcp2l1 and Klf2 provide a coordinated balance, with concurrent yet opposing inputs to Oct4 gene regulation.

In future studies, it should be determined whether these interactions are direct, and whether they are of functional significance for ESC equilibrium. This may require the implementation of quantitative (not Boolean) computational models and rapid knock-out technologies.

8.2.3 Does Klf2 play a post-implantation role in mesodermal lineage specification?

Paradoxically, the ability of Klf2 to reprogram EpiSCs to naïve pluripotency appears secondary to its ability to upregulate mesodermal markers (Chapter 5). I did not observe naïve gene activation until ≥48 h after Klf2 induction, whereas mesodermal markers were strongly induced. The investigation in Chapter 7 led me to conclude that the ability of Klf2 to drive reprogramming lies largely in its ability to maintain Oct4 expression, endowing a prolonged window of opportunity for 2iLIF signals to eventually induce naïve pluripotency. Interestingly, Oct4 is also maintained during the first stages of mesendoderm differentiation (Downs, 2008; Thomson et al., 2011). I wonder if these two observations are not coincidental, but rather that by inducing a mesodermal Oct4-expressing program, Klf2 serendipitously provides the opportunity for the environment to drive reprogramming.

With these ideas in mind, I turned back to the literature to search for evidence as to whether Klf2 is re-expressed again shortly after implantation, and whether it might play a role in regulation of mesodermal lineage specification *in vivo*. An early study found Klf2 transcript at E7.5 (Anderson et al., 1995). More recently, Klf2 expression has been detected at E8.5 in endocardial and endothelial compartments, increasing to E10.5 in the heart tube loops and endocardial cells (Lee et al., 2006). *Klf2-/-* mice die between E11.5–14.5 (Chiplunkar et al., 2013; Kuo et al., 1997). This embryonic lethality indicates an important post-implantation role for Klf2 that is not compensated by other Klf family members, and has been attributed to either haemorrhage or embryonic heart failure (Chiplunkar et al., 2013; Kuo et al., 1997; Lee et al., 2006). Intriguingly, there is evidence implicating Klf2 as a regulator of endothelial-to-mesenchymal transformation in the atrioventricular canal prior to E10.5 (Chiplunkar et al., 2013).

In established ESCs, Klf2 clearly has a positive role in maintaining the naïve identity (Hall et al., 2009; Jiang et al., 2008; Niwa et al., 2009; Yamane et al., 2018; Yeo et al., 2014). Yet, as

with many TFs governing naïve pluripotency, Klf2 also has expressions and roles elsewhere in development (Bialkowska et al., 2017), including early post-implantation. From the point of view of an EpiSC, a cell type on the cusp of gastrulation, why would it respond to such a factor according to its role in the earlier naïve network rather according to forwards developmental logic?

In a future project, the biological questions that I would suggest to ask are as follows: does Klf2 actively specify particular lineages as the post-implantation epiblast progresses? In the absence of instructive signals, for how long are intermediates from iKlf2 EpiSCs open to both naïve and post-implantation fates? What is the limit after which mesodermal induction hinders rather than helps reprogramming? How are different roles of Klf2 computed in different contexts, and do they mechanistically relate? What is it about the pre-implantation vs post-implantation context that can switch which program Klf2 directs? These questions could be initially explored *in vitro* by differentiation assays in neural, mesodermal or 'open' differentiation conditions, starting from iKlf2 EpiSCs and/or iPSCs ± dox[9], and *Klf2-/-* cells. Subsequently, behaviour of labelled *Klf2-/-* and iKlf2 ESCs/EpiSCs could be traced in chimeric embryos or gastruloids. Since EpiSCs do not express Klf2 themselves, yet temporally are positioned between Klf2-expressing naïve and post-implantation tissues, this could provide an excellent paradigm for the study of factor reuse with context-dependent interpretive logics.

8.2.4 Sox2 mRNA regulation and Oct/Sox partner swapping

In counterpoint to its robust support of Oct4 expression, Klf2 silences Sox2 within one hour post-induction (Figure 5.2.1A). Since the half-life of Sox2 mRNA is 85-120 min (Ormsbee Golden et al., 2013; Sharova et al., 2009), this suggests active mRNA degradation rather than simply transcriptional suppression. In human mesenchymal stem cells, there is evidence for Sox2 mRNA degradation mediated by the 3' UTR RNA-binding protein ELAV1 together with RISC and miR145 (Latorre et al., 2016). miR145 has also been implicated in the suppression of Sox2 in human ESCs and murine neural stem cells (Morgado et al.; Xu et al., 2009). Unfortunately miR expressions cannot be analysed in our RNAseq data due to size-exclusion from the libraries, but this would be worth pursuing in the future. In a preliminary experiment I found that forced Sox2 expression abolished iKlf2-driven reprogramming, suggesting that transient Sox2 silencing is essential for this transition route. Since constitutive Sox2 was driven

[9] In a preliminary experiment, iKlf2 cells +dox dominantly differentiated into a 'motile mesoderm' morphology, even in strongly pro-neural N2B27+A8301 conditions.

from a CAG promoter and lacked the endogenous 3' UTR, further work will be required to distinguish between the functional importance of transcriptional vs post-transcriptional silencing, and the mechanisms thereof.

Pou and Sox family proteins usually bind DNA in cooperation with partner factors. This partnering is required for transcriptional regulation, and different partnerships bind different gene targets. There is extensive precedent for Pou/Sox partner swapping as a critical regulator of lineage specification (Aksoy et al., 2013; Kondoh and Kamachi, 2010). I wonder whether Sox2 silencing reflects a mechanism by which Klf2 influences identity transitions, by altering Oct4 partnering. Several Pou/Sox family members are rapidly modulated following Klf2 induction in EpiSCs, as is Oct4-partner Otx2 (Figure 8.2.1). In the first instance, these transcriptional trends should be confirmed on the protein level. Then, during both reprogramming and differentiation of iKlf2 EpiSCs, Oct4 immunoprecipitation could ascertain whether its partners do indeed swap from Oct4/Sox2 and Oct4/Otx2 to Oct4/Sox7 or other factors. ChIP studies could then evaluate how this modulates Oct4 target genes.

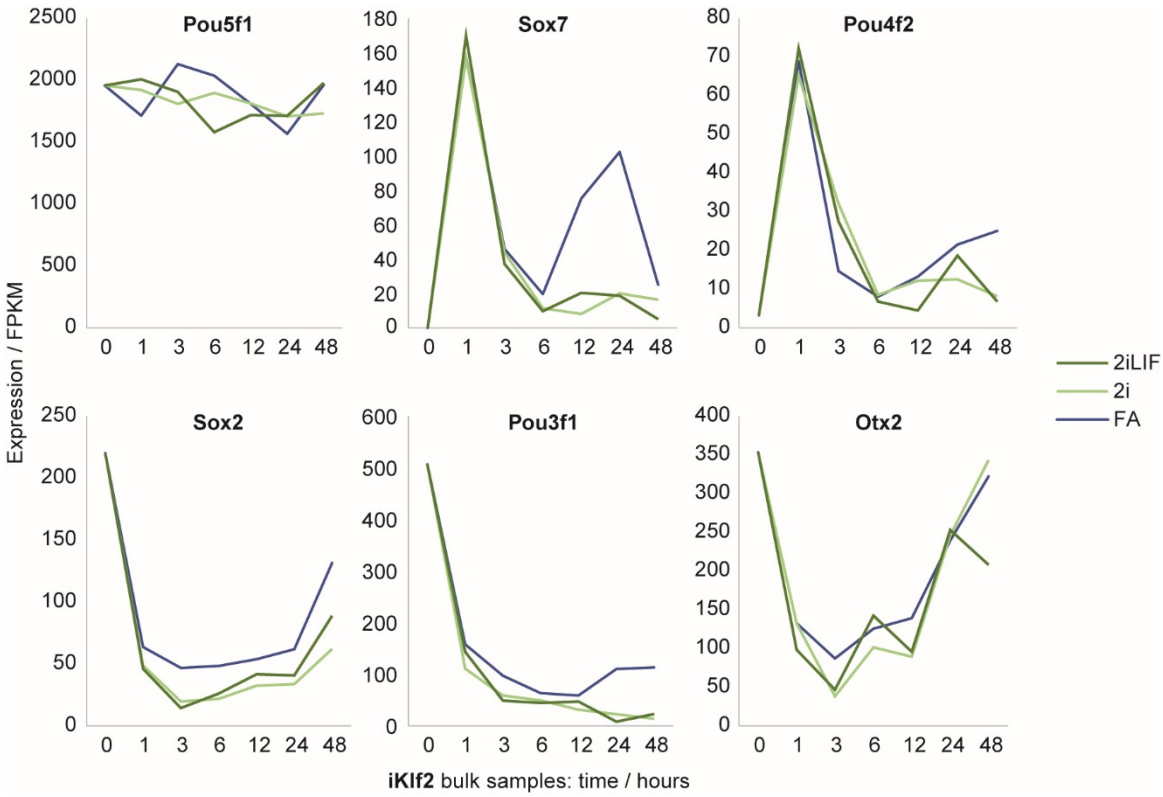


Figure 8.2.1: Pou/Sox family responses to Klf2 induction in EpiSCs

Timecourse gene expression analysis of iKlf2 EpiSCs induced with dox in the indicated conditions. Gene expressions are presented as FPKM from RNAseq.

8.2.5 Transition factors during regenerative fate transitions *in vivo*?

In this thesis, I found that appropriate expression of Oct4 is the pivotal feature for transition into the naïve pluripotent identity. Indeed, maintenance of this transition factor appears to be more important than the transcriptional program directly induced by the reprogramming driver, as exemplified by the case of Klf2 where induction of mesodermal rather than naïve genes is ultimately compatible with efficient reprogramming, thanks to the interim robust support of Oct4 expression. Does this concept extrapolate to establishment of other cellular identities? Are there other cases whereby multiple routes of identity change converge on a unifying transition factor? Intuitively, this could provide benefits when responding to unpredictable events such as injury, balancing flexibility with fidelity of cell identity regeneration.

Compared to the tractable *in vitro* systems employed in this thesis, answering these questions in a variable-controlled manner during *in vivo* identity transitions will be far more challenging. Ongoing advances in our understanding and manipulation of regenerative events may open this avenue in the near future. For example, induction of progenitors in response to liver damage is increasingly well defined thanks to the published and ongoing work of Meri Huch (Huch et al., 2013). In response to Wnt signalling, differentiated duct cells undergo dedifferentiation and epigenetic reprogramming to yield proliferative Lgr5+ bipotent liver progenitors. The ability of single Lgr5+ progenitors to form *in vitro* liver organoids provides a clear confirmation of destination molecular signature and functionality. Therefore, it could now be feasible to study fate transition into the liver progenitor identity from a variety of starting liver positions or damage types.

Recent work on Axolotl limb regeneration has also highlighted and clarified a process of tissue dedifferentiation (Tanaka group, in press). During this reprogramming there is a marked reduction in cellular heterogeneity, indicative of a transient convergence of cellular identities prior to redifferentiation. Does a transition factor regulate entry to this putative attractor state? What determines the relative ease with which the Axolotl can re-access this? Thanks to genomic sequencing and molecular tools for Axolotl manipulation (Nowoshilow et al., 2018), we are now in a position to tackle these fascinating questions.

As a community, I hope we will continue to identify the transition factors and supporting logic for the multitude of developmental, regenerative and pathological cell identity transitions.

CHAPTER 9: BIBLIOGRAPHY

Adachi, K., Nikaido, I., Ohta, H., Ohtsuka, S., Ura, H., Kadota, M., Wakayama, T., Ueda, H.R., and Niwa, H. (2013). Context-dependent wiring of Sox2 regulatory networks for self-renewal of embryonic and trophoblast stem cells. Mol. Cell *52*, 380–392.

Adachi, K., Kopp, W., Wu, G., Heising, S., Greber, B., Stehling, M., Araúzo-Bravo, M.J., Boerno, S.T., Timmermann, B., Vingron, M., et al. (2018). Esrrb Unlocks Silenced Enhancers for Reprogramming to Naive Pluripotency. Cell Stem Cell.

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., and Bazett-Jones, D.P. (2010). Global Chromatin Architecture Reflects Pluripotency and Lineage Commitment in the Early Mouse Embryo. PLoS One *5*, e10531.

Aksoy, I., Jauch, R., Chen, J., Dyla, M., Divakar, U., Bogu, G.K., Teo, R., Leng Ng, C.K., Herath, W., Lili, S., et al. (2013). Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. EMBO J. *32*, 938–953.

Ambrosetti, D.C., Basilico, C., and Dailey, L. (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol. Cell. Biol. *17*, 6321–6329.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Anderson, K.G. V., Hamilton, W.B., Roske, F. V., Azad, A., Knudsen, T.E., Canham, M.A., Forrester, L.M., and Brickman, J.M. (2017). Insulin fine-tunes self-renewal pathways governing naive pluripotency and extra-embryonic endoderm. Nat. Cell Biol.

Anderson, K.P., Kern, C.B., Crable, S.C., and Lingrel, J.B. (1995). Isolation of a Gene Encoding a Functional Zinc Finger Protein Homologous to Erythroid Krüppel-Like Factor: Identification of a New Multigene Family. Mol. Cell. Biol. *15*, 5957–5965.

Auclair, G., Guibert, S., Bender, A., and Weber, M. (2014). Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. Genome Biol. *15*, 545.

Auernhammer, C.J., Bousquet, C., and Melmed, S. (1999). Autoregulation of pituitary corticotroph SOCS-3 expression: characterization of the murine SOCS-3 promoter. Proc Natl Acad Sci U S A *96*, 6964–6969.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. *8*, 532–538.

Beddington, R.S., and Robertson, E.J. (1999). Axis development and early asymmetry in mammals. Cell *96*, 195–209.

van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369–381.

Bernemann, C., Greber, B., Ko, K., Sterneckert, J., Han, D.W., Araúzo-Bravo, M.J., and Schöler, H.R. (2011). Distinct Developmental Ground States of Epiblast Stem Cell Lines Determine Different Pluripotency Features. Stem Cells *29*, 1496–1503.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bhattacharya, D., Talwar, S., Mazumder, A., and Shivashankar, G. V (2009). Spatio-temporal plasticity in chromatin organization in mouse cell differentiation and during Drosophila embryogenesis. Biophys. J. *96*, 3832–3839.

Bialkowska, A.B., Yang, V.W., and Mallipattu, S.K. (2017). Krüppel-like factors in mammalian stem cells and development. Development *144*, 737–754.

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature *500*, 222–226.

Blauwkamp, T.A., Nigam, S., Ardehali, R., Weissman, I.L., and Nusse, R. (2012). Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. Nat. Commun. *3*, 1070.

Boeuf, H., Hauss, C., Graeve, F.D., Baran, N., and Kedinger, C. (1997). Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. J Cell Biol *138*, 1207–1217.

Böhmer, F.-D., and Friedrich, K. (2014). Protein tyrosine phosphatases as wardens of STAT signaling. JAK-STAT *3*, e28087.

Bonni, A., Sun, Y., Nadal-Vicens, M., Bhatt, A., Frank, D.A., Rozovsky, I., Stahl, N., Yancopoulos, G.D., and Greenberg, M.E. (1997). Regulation of gliogenesis in the central nervous system by the JAK-STAT signaling pathway. Science *278*, 477–483.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat. Cell Biol. *16*, 516–528.

Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A., and Bertone, P. (2015). Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. Dev. Cell *35*, 366–382.

Bourillot, P.Y., Aksoy, I., Schreiber, V., Wianny, F., Schulz, H., Hummel, O., Hubner, N., and Savatier, P. (2009). Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. Stem Cells *27*.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255–256.

Briscoe, J., and Small, S. (2015). Morphogen rules: design principles of gradient-mediated embryo patterning. Development *142*, 3996–4009.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191–195.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., Van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*.

Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Dev. Biol. *210*, 30–43.

Carbognin, E., Betto, R.M., Soriano, M.E., Smith, A.G., and Martello, G. (2016). Stat3 promotes mitochondrial transcription and oxidative respiration during maintenance and induction of naive pluripotency. EMBO J. *35*, 618–634.

Cavallo, R.A., Cox, R.T., Moline, M.M., Roose, J., Polevoy, G.A., Clevers, H., Peifer, M., and Bejsovec, A. (1998). Drosophila Tcf and Groucho interact to repress Wingless signalling activity. Nature *395*, 604–608.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. Cell *113*, 643–655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. Dev. Cell *10*, 615–624.

Chen, J., Liu, J., Yang, J., Chen, Y., Chen, J., Ni, S., Song, H., Zeng, L., Ding, K., and Pei, D. (2011). BMPs functionally replace Klf4 and support efficient reprogramming of mouse fibroblasts by Oct4 alone. Cell Res. *21*, 205–212.

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Wu, Y., Guo, L., et al. (2013). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat Genet *45*, 34–42.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B.V.B., Wong, E., Orlov, Y.L.L.Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Chiplunkar, A.R., Lung, T.K., Alhashem, Y., Koppenhaver, B.A., Salloum, F.N., Kukreja, R.C., Haar, J.L., and Lloyd, J.A. (2013). Krüppel-Like Factor 2 Is Required for Normal Mouse Cardiac Development. PLoS One *8*, e54891.

Choi, J., Huebner, A.J., Clement, K., Walsh, R.M., Savol, A., Lin, K., Gu, H., Di Stefano, B., Brumbaugh, J., Kim, S.-Y., et al. (2017). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. Nature *548*, 219.

Clavel, C., Grisanti, L., Zemla, R., Rezza, A., Barros, R., Sennett, R., Mazloom, A.R., Chung, C.-Y., Cai, X., Cai, C.-L., et al. (2012). Sox2 in the Dermal Papilla Niche Controls Hair Growth by Fine-Tuning BMP Signaling in Differentiating Hair Shaft Progenitors. Dev. Cell *23*, 981–994.

Clevers, H. (2006). Wnt/beta-catenin signaling in development and disease. Cell *127*, 469–480.

Corsinotti, A., Wong, F.C., Tatar, T., Szczerbinska, I., Halbritter, F., Colby, D., Gogolok, S., Pantier, R., Liggat, K., Mirfazeli, E.S., et al. (2017). Distinct SoxB1 networks are required for naïve and primed pluripotency. Elife *6*.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P. V, Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370–374.

Cuny, G.D., Yu, P.B., Laha, J.K., Xing, X., Liu, J.-F., Lai, C.S., Deng, D.Y., Sachidanandan, C., Bloch, K.D., and Peterson, R.T. (2008). Structure–activity relationship study of bone morphogenetic protein (BMP) signaling inhibitors. Bioorg. Med. Chem. Lett. *18*, 4388–4392.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002). A genomic regulatory network for development. Science *295*, 1669–1678.

Davidson, K.C., Mason, E.A., and Pera, M.F. (2015). The pluripotent state in mouse and human. *142*, 3090–3099.

Davila-Velderrain, J., Martinez-Garcia, J.C., and Alvarez-Buylla, E.R. (2015). Modeling the epigenetic attractors landscape: toward a post-genomic mechanistic understanding of development. Front. Genet. *6*, 160.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science (80-.). *343*, 193–196.

DeVeale, B., Brokhman, I., Mohseni, P., Babak, T., Yoon, C., Lin, A., Onishi, K., Tomilin, A., Pevny, L., Zandstra, P.W., et al. (2013). Oct4 Is Required ~E7.5 for Proliferation in the Primitive Streak. PLoS Genet. *9*, e1003957.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Do, D. V., Ueda, J., Messerschmidt, D.M., Lorthongpanich, C., Zhou, Y., Feng, B., Guo, G., Lin, P.J., Hossain, M.Z., Zhang, W., et al. (2013). A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. Genes Dev. *27*, 1378–1390.

Doble, B.W., and Woodgett, J.R. (2003). GSK-3: tricks of the trade for a multi-tasking kinase. J. Cell Sci. *116*, 1175–1186.

Downs, K.M. (2008). Systematic localization of oct-3/4 to the gastrulating mouse conceptus suggests manifold roles in mammalian development. Dev. Dyn. *237*, 464–475.

Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics *24*, 1547–1548.

Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science (80-.). *344*.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D.G., Buetow, K.H., et al. (2008). Global Transcription in Pluripotent Embryonic Stem Cells. Cell Stem Cell *2*, 437–447.

Egwuagu, C.E. (2009). STAT3 in CD4+ T helper cell differentiation and inflammatory diseases. Cytokine *47*, 149–156.

Enver, T., Pera, M., Peterson, C., Andrews, P.W., Doble, B., Woodgett, J., Cohen, P., Smith, A., Panchision, D.M., McKay, R.D., et al. (2009). Stem Cell States, Fates, and the Rules of Attraction. Cell Stem Cell *4*, 387–397.

Esch, D., Vahokoski, J., Groves, M.R., Pogenberg, V., Cojocaru, V., Vom Bruch, H., Han, D., Drexler, H.C.A., Araúzo-Bravo, M.J., Ng, C.K.L., et al. (2013). A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat. Cell Biol. *15*, 295–301.

Esteban, M.A., and Pei, D. (2012). Vitamin C improves the quality of somatic cell reprogramming. Nat Genet *44*, 366–367.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Ezashi, T., Yuan, Y., and Roberts, R.M. (2016). Pluripotent Stem Cells from Domesticated Mammals. Annu. Rev. Anim. Biosci *4*, 223–253.

Factor, D.C.C., Corradin, O., Zentner, G.E.E., Saiakhova, A., Song, L., Chenoweth, J.G.G., McKay, R.D.D., Crawford, G.E.E., Scacheri, P.C.C., and Tesar, P.J.J. (2014). Epigenomic comparison reveals activation of “seed” enhancers during transition from naive to primed pluripotency. Cell Stem Cell *14*, 854–863.

Farah, M.H., Olson, J.M., Sucic, H.B., Hume, R.I., Tapscott, S.J., and Turner, D.L. (2000). Generation of neurons by transient expression of neural bHLH proteins in mammalian cells. Development *127*, 693–702.

Ferrell, J.E. (2012). Bistability, bifurcations, and Waddington's epigenetic landscape. Curr. Biol. *22*, R458-66.

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell *11*, 477–490.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., et al. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell *13*, 351–359.

Fidalgo, M., Faiola, F., Pereira, C.F., Ding, J., Saunders, A., Gingold, J., Schaniel, C., Lemischka, I.R., Silva, J.C., and Wang, J. (2012). Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. Proc Natl Acad Sci U S A *109*, 16202–16207.

Friedrich Ben-Nun, I., Montague, S.C., Houck, M.L., Tran, H.T., Garitaonandia, I., Leonardo, T.R., Wang, Y.-C., Charter, S.J., Laurent, L.C., Ryder, O.A., et al. (2011). Induced pluripotent stem cells from highly endangered species. Nat. Methods *8*, 829–831.

Futschik, M.E., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. J. Bioinform. Comput. Biol. *3*, 965–988.

Gagliardi, A., Mullin, N.P., Ying Tan, Z., Colby, D., Kousa, A.I., Halbritter, F., Weiss, J.T., Felker, A., Bezstarosti, K., Favaro, R., et al. (2013). A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. EMBO J. *32*, 2231–2247.

Galonska, C., Ziller, M.J., Karnik, R., and Meissner, A. (2015). Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. Cell Stem Cell *17*, 462–470.

Gardner, R.L., and Lyon, M.F. (1971). Biological sciences: X chromosome inactivation studied by injection of a single cell into the mouse blastocyst. Nature *231*, 385–386.

Gardner, R.L., and Rossant, J. (1979). Investigation of the fate of 4-5 day post-coitum mouse inner cell mass cells by blastocyst injection. J. Embryol. Exp. Morphol. *52*, 141–152.

Gardner, R.L., Lyon, M.F., Evans, E.P., and Burtenshaw, M.D. (1985). Clonal analysis of X-chromosome inactivation and the origin of the germ line in the mouse embryo. J. Embryol. Exp. Morphol. *88*, 349–363.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., McManus, M.T., Plath, K., Meshorer, E., et al. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature *460*, 863–868.

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. Nat. Rev. Mol. Cell Biol. *12*, 36–47.

Gaspar-Maia, A., Qadeer, Z.A., Hasson, D., Ratnakumar, K., Adrian Leu, N., Leroy, G., Liu, S., Costanzi, C., Valle-Garcia, D., Schaniel, C., et al. (2013). MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. Nat. Commun. *4*, 1565.

Gillich, A., Bao, S., Grabole, N., Hayashi, K., Trotter, M.W.B.M.W.M.W.B., Pasque, V., Magnúsdóttir, E., Surani, M.A.A.M.A., Agrelo, R., Souabni, A., et al. (2012). Epiblast Stem Cell-Based System Reveals Reprogramming Synergy of Germline Factors. Cell Stem Cell *10*, 425–439.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. Cell *128*, 635–638.

Gong, T., and Szustakowski, J.D. (2013). DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. Bioinformatics *29*, 1083–1085.

Gontan, C., de Munck, A., Vermeij, M., Grosveld, F., Tibboel, D., and Rottier, R. (2008). Sox2 is important for two crucial processes in lung development: Branching morphogenesis and epithelial cell differentiation. Dev. Biol. *317*, 296–309.

Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science (80-.). *268*, 1766–1769.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Graham, S.J.L., Wicher, K.B., Jedrusik, A., Guo, G., Herath, W., Robson, P., and Zernicka-Goetz, M. (2014). BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. Nat. Commun. *5*, 5667.

Greber, B., Wu, G., Bernemann, C., Joo, J.Y., Han, D.W., Ko, K., Tapia, N., Sabour, D., Sterneckert, J., Tesar, P., et al. (2010). Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. Cell Stem Cell *6*, 215–226.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development *136*, 1063–1069.

Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. Dev. Cell *18*, 675–685.

Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., and Nichols, J. (2016). Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. Stem Cell Reports *6*, 437–446.

Guo, G., von Meyenn, F., Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W., et al. (2017). Epigenetic resetting of human pluripotency. Development *144*.

Gurdon, J.B. (1962). The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles. Development *10*.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Habibi, E., Brinkman, A.B.B., Arand, J., Kroeze, L.I.I., Kerstens, H.H.D.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C.C., et al. (2013). Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. Cell Stem Cell *13*, 360–369.

Hackett, J.A., and Surani, M.A. (2014). Regulatory Principles of Pluripotency: From the Ground State Up. Cell Stem Cell *15*, 416–430.

Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., and Surani, M.A. (2013). Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. Stem Cell Reports *1*, 518–531.

Hackett, J.A., Kobayashi, T., Dietmann, S., and Surani, M.A. (2017). Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum. Stem Cell Reports *8*, 1645–1658.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech. Dev. *117*.

Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M.A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature *452*, 877–881.

Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R., et al. (2009). Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. Cell Stem Cell *5*, 597–609.

Han, D.W., Tapia, N., Joo, J.Y., Greber, B., Araúzo-Bravo, M.J., Bernemann, C., Ko, K., Wu, G., Stehling, M., Do, J.T., et al. (2010). Epiblast Stem Cell Subpopulations Represent Mouse Embryos of Distinct Pregastrulation Stages. Cell *143*, 617–627.

Hao, J., Ho, J.N., Lewis, J.A., Karim, K.A., Daniels, R.N., Gentry, P.R., Hopkins, C.R., Lindsley, C.W., and Hong, C.C. (2010). In vivo structure - Activity relationship study of dorsomorphin analogues identifies selective VEGF and BMP inhibitors. ACS Chem. Biol. *5*, 245–253.

Hart, A.H., Hartley, L., Ibrahim, M., and Robb, L. (2004). Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. Dev. Dyn. *230*, 187–198.

Hassan-Zadeh, V., Rugg-Gunn, P., and Bazett-Jones, D.P. (2017). DNA methylation is dispensable for changes in global chromatin architecture but required for chromocentre formation in early stem cell differentiation. Chromosoma *126*, 605–614.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells. Cell Stem Cell *6*, 479–491.

Hayashi, K., Lopes, S.M.C. de S., Tang, F., and Surani, M.A. (2008). Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. Cell Stem Cell *3*, 391–401.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. Cell *146*.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Herrera-Delgado, E., Perez-Carrasco, R., Briscoe, J., and Sollich, P. (2018). Memory functions reveal structural properties of gene regulatory networks. PLOS Comput. Biol. *14*, e1006003.

Ho, J.W.K., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.-A., Minoda, A., Tolstorukov, M.Y., Appert, A., et al. (2014). Comparative analysis of metazoan chromatin organization. Nature *512*, 449–452.

Hough, S.R., Thornton, M., Mason, E., Mar, J.C., Wells, C.A., and Pera, M.F. (2014). Single-Cell Gene Expression Profiles Define Self-Renewing, Pluripotent, and Lineage Primed States of Human Pluripotent Stem Cells. Stem Cell Reports *2*, 881–895.

Huang, S. (2012). The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology? BioEssays *34*, 149–157.

Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D.E. (2005). Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network. Phys. Rev. Lett. *94*, 128701.

Huang, S., Guo, Y.-P., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Genomes Dev. Control.

Huang, Y., Osorno, R., Tsakiridis, A., and Wilson, V. (2012). In Vivo Differentiation Potential of Epiblast Stem Cells Revealed by Chimeric Embryo Formation. Cell Rep. *2*, 1571–1578.

Huch, M., Dorrell, C., Boj, S.F., van Es, J.H., Li, V.S.W., van de Wetering, M., Sato, T., Hamer, K., Sasaki, N., Finegold, M.J., et al. (2013). In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. Nature *494*, 247–250.

Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., Birchmeier, C., and Birchmeier, W. (2000). Requirement for beta-catenin in anterior-posterior axis formation in mice. J. Cell Biol. *148*, 567–578.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature *442*, 533–538.

Iwafuchi-Doi, M., Matsuda, K., Murakami, K., Niwa, H., Tesar, P.J., Aruga, J., Matsuo, I., and Kondoh, H. (2012). Transcriptional regulatory networks in epiblast cells and during anterior neural plate development as modeled in epiblast stem cells. Development *139*, 3926–3937.

Jaeger, J., and Monk, N. (2014). Bioattractors: dynamical systems theory and the evolution of regulatory processes. J. Physiol. *592*, 2267–2281.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol *10*, 353–360.

Johnston, P.A., and Grandis, J.R. (2011). STAT3 signaling: anticancer strategies and challenges. Mol Interv *11*, 18–26.

Jost, D. (2014). Bifurcation in epigenetics: Implications in development, proliferation, and diseases. Phys. Rev. E *89*, 010701.

Juliá, M., Telenti, A., and Rausell, A. (2015). Sincell: an R/Bioconductor package for statistical assessment of cell-state hierarchies from single-cell RNA-seq. Bioinformatics *31*, 3380–3382.

Kahan, B.W., and Ephrussi, B. (1970). Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. J. Natl. Cancer Inst. *44*, 1015–1036.

Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H.J., Nett, I., Marks, H., Walker, R., Stunnenberg, H.G., Lilley, K.S., et al. (2017). Tracking the embryonic stem cell transition from ground state pluripotency. Development *144*, 1221–1234.

Karwacki-Neisius, V., Göke, J., Osorno, R., Halbritter, F., Ng, J.H., Weiße, A.Y., Wong, F.C.K., Gagliardi, A., Mullin, N.P., Festuccia, N., et al. (2013). Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. Cell Stem Cell *12*, 531–545.

Kauffman, S.A. (1993). The Origins of Order: self-organisation and selection in evolution.

Kaufman, M.H. (1992). The atlas of mouse development (Academic Press).

Keramari, M., Razavi, J., Ingman, K.A., Patsch, C., Edenhofer, F., Ward, C.M., and Kimber, S.J. (2010). Sox2 Is Essential for Formation of Trophectoderm in the Preimplantation Embryo. PLoS One *5*, e13952.

Kershaw, N.J., Murphy, J.M., Liau, N.P., Varghese, L.N., Laktyushin, A., Whitlock, E.L., Lucet, I.S., Nicola, N.A., and Babon, J.J. (2013). SOCS3 binds specific receptor-JAK complexes to control cytokine signaling by direct kinase inhibition. Nat Struct Mol Biol *20*, 469–476.

Kim, H., Wu, J., Ye, S., Tai, C.-I., Zhou, X., Yan, H., Li, P., Pera, M., and Ying, Q.-L. (2013). Modulation of β-catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. Nat. Commun. *4*, 2403.

Kim, J.B., Sebastiano, V., Wu, G., Araúzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D., et al. (2009). Oct4-Induced Pluripotency in Adult Neural Stem Cells. Cell *136*, 411–419.

King, H.W., and Klose, R.J. (2017). The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. Elife *6*, e22631.

Kisseleva, T., Bhattacharya, S., Braunstein, J., and Schindler, C.W. (2002). Signaling through the JAK/STAT pathway, recent advances and future challenges. Gene *285*, 1–24.

Kojima, Y., Kaufman-Francis, K., Studdert, J.B.B., Steiner, K.A.A., Power, M.D.D., Loebel, D.A.F.A.F., Jones, V., Hor, A., de Alencastro, G., Logan, G.J.J., et al. (2014). The Transcriptional and Functional Properties of Mouse Epiblast Stem Cells Resemble the Anterior Primitive Streak. Cell Stem Cell *14*, 107–120.

Kondoh, H., and Kamachi, Y. (2010). SOX–partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. Int. J. Biochem. Cell Biol. *42*, 391–399.

Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells *26*, 903–911.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285–294.

Krejčí, J., Uhlířová, R., Galiová, G., Kozubek, S., Šmigová, J., and Bártová, E. (2009). Genome-wide reduction in H3K9 acetylation during human embryonic stem cell differentiation. J. Cell. Physiol. *219*, 677–687.

Krijthe, J.H. (2015). Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation.

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development *134*, 2895–2902.

Kuo, C.T., Veselits, M.L., Barton, K.P., Lu, M.M., Clendenin, C., and Leiden, J.M. (1997). The LKLF transcription factor is required for normal tunica media formation and blood vessel stabilization during murine embryogenesis. Genes Dev. *11*, 2996–3006.

Kurek, D., Neagu, A., Tastemel, M., Tüysüz, N., Lehmann, J., van de Werken, H.J.G., Philipsen, S., van der Linden, R., Maas, A., van IJcken, W.F.J., et al. (2015). Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. Stem Cell Reports *4*, 114–128.

Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S. -y., Suemori, H., Nakatsuji, N., and Tada, T. (2005). Octamer and Sox Elements Are Required for Transcriptional cis Regulation of Nanog Gene Expression. Mol. Cell. Biol. *25*, 2475–2485.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497–501.

Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., et al. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature *498*, 511–515.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Lanner, F., Lee, K.L., Sohl, M., Holmborn, K., Yang, H., Wilbertz, J., Poellinger, L., Rossant, J., and Farnebo, F. (2010). Heparan Sulfation Dependent FGF Signalling Maintains Embryonic Stem Cells Primed for Differentiation in a Heterogeneous State. Stem Cells *28*, N/A-N/A.

Latorre, E., Carelli, S., Caremoli, F., Giallongo, T., Colli, M., Canazza, A., Provenzani, A., Di Giulio, A.M., and Gorio, A. (2016). Human Antigen R Binding and Regulation of SOX2 mRNA in Human Mesenchymal Stem Cells. Mol. Pharmacol. *89*, 243–252.

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. J. Stat. Softw. *25*, 1–18.

Lee, J.S., Yu, Q., Shin, J.T., Sebzda, E., Bertozzi, C., Chen, M., Mericko, P., Stadtfeld, M., Zhou, D., Cheng, L., et al. (2006). Klf2 is an essential regulator of vascular hemodynamic forces in vivo. Dev. Cell *11*, 845–857.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799–804.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316.

Levine, M., and Davidson, E.H. (2005). Gene regulatory networks for development. Proc. Natl. Acad. Sci. U. S. A. *102*, 4936–4942.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516–520.

Lin, S.M., Du, P., Huber, W., and Kibbe, W.A. (2008). Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic Acids Res *36*, e11.

Liu, P., Wakamiya, M., Shea, M.J., Albrecht, U., Behringer, R.R., and Bradley, A. (1999). Requirement for Wnt3 in vertebrate axis formation. Nat. Genet. *22*, 361–365.

Liu, P., Chen, M., Liu, Y., Qi, L.S., and Ding, S. (2018). CRISPR-Based Chromatin Remodeling of the Endogenous Oct4 or Sox2 Locus Enables Reprogramming to Pluripotency. Cell Stem Cell.

Loh, K.M., and Lim, B. (2011). A precarious balance: pluripotency factors as lineage specifiers. Cell Stem Cell *8*, 363–369.

Loh, Y.-H., Zhang, W., Chen, X., George, J., and Ng, H.-H. (2007). Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. Genes Dev. *21*.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431–440.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lyashenko, N., Winter, M., Migliorini, D., Biechele, T., Moon, R.T., and Hartmann, C. (2011). Differential requirement for the dual functions of β-catenin in embryonic stem cell self-renewal and germ layer formation. Nat. Cell Biol. *13*, 753–761.

MacArthur, B.D., Sevilla, A., Lenz, M., Müller, F.-J., Schuldt, B.M., Schuppert, A.A., Ridden, S.J., Stumpf, P.S., Fidalgo, M., Ma'ayan, A., et al. (2012). Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. Nat. Cell Biol. *14*, 1139–1147.

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., and Brockdorff, N. (2004). Reactivation of the Paternal X Chromosome in Early Mouse Embryos. Science (80-.). *303*, 666–669.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590–604.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Martello, G., and Smith, A. (2014). The Nature of Embryonic Stem Cells. Annu. Rev. Cell Dev. Biol. *30*, 647–675.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell *11*, 491–504.

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J *32*, 2561–2574.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. *78*, 7634–7638.

Martin, G.R., and Evans, M.J. (1975). Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. Proc. Natl. Acad. Sci. U. S. A. *72*, 1441–1445.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat. Cell Biol. *9*, 625–635.

Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., and Yokota, T. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. EMBO J *18*, 4261–4269.

Merrill, B.J., Pasolli, H.A., Polak, L., Rendl, M., García-García, M.J., Anderson, K. V, and Fuchs, E. (2004). Tcf3: a transcriptional regulator of axis induction in the early embryo. Development *131*, 263–274.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells. Dev. Cell *10*, 105–116.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49–55.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. *20*, 1215–1228.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development *99*.

Morgado, A.L., Ilia, C., Rodrigues, M.P., and So A, S. MicroRNA-145 Regulates Neural Stem Cell Differentiation Through the Sox2–Lin28/let-7 Signaling Pathway.

Morgani, S.M., and Brickman, J.M. (2015). LIF supports primitive endoderm expansion during pre-implantation development. Development *142*, 3488–3499.

Morgani, S., Nichols, J., and Hadjantonakis, A.-K. (2017). The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. BMC Dev. Biol. *17*, 7.

Moris, N., Pina, C., and Arias, A.M. (2016). Transition states and cell fate decisions in epigenetic landscapes. Nat. Rev. Genet. *17*, 693–703.

Mousavi, K., Zare, H., Dell'Orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. Mol. Cell *51*, 606–617.

Murakami, K., Günesdogan, U., Zylicz, J.J., Tang, W.W.C.C., Sengupta, R., Kobayashi, T., Kim, S., Butler, R., Dietmann, S., Surani, M.A., et al. (2016). NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. Nature *529*, 403–407.

Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., Yagi, K., Miyazaki, J. -i., Matoba, R., Ko, M.S.H., et al. (2006). Klf4 Cooperates with Oct3/4 and Sox2 To Activate the Lefty1 Core Promoter in Embryonic Stem Cells. Mol. Cell. Biol. *26*, 7772–7782.

Nashun, B., Hill, P.W.S., and Hajkova, P. (2015). Reprogramming of cell fate: epigenetic memory and the erasure of memories past. EMBO J. *34*, 1296–1308.

Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Göttgens, B. (2016). A single cell resolution map of mouse haematopoietic stem and progenitor cell differentiation. Blood.

Nichols, J., and Smith, A. (2009). Naive and Primed Pluripotent States. Cell Stem Cell *4*, 487–492.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., Smith, A., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379–391.

Nichols, J., Chambers, I., Taga, T., and Smith, A. (2001). Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. Development *128*, 2333–2339.

Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development *136*, 3215–3222.

Nishimoto, M., Fukushima, A., Okuda, A., and Muramatsu, M. (1999). The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. Mol. Cell. Biol. *19*, 5453–5465.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat. Methods *6*, 917–922.

Nishiyama, A., Sharov, A.A., Piao, Y., Amano, M., Amano, T., Hoang, H.G., Binder, B.Y., Tapnio, R., Bassey, U., Malinou, J.N., et al. (2013). Systematic repression of transcription factors reveals limited patterns of gene expression changes in ES cells. Sci. Rep. *3*, 1390.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev *12*, 2048–2060.

Niwa, H., Miyazaki, J., Smith, A.G., and Miyazaki, J. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118–122.

Nowoshilow, S., Schloissnig, S., Fei, J.-F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature *554*, 50–55.

O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., Kaji, K., O'Malley, J., et al. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature *499*, 88–91.

Ogorevc, J., Orehek, S., and Dovč, P. (2016). Cellular reprogramming in farm animals: an overview of iPSC generation in the mammalian farm animal species. J. Anim. Sci. Biotechnol.

Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, J.M., Guo, T., Qi, Z., Downey, S.L., Manos, P.D., Rossi, D.J., et al. (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat. Cell Biol. *13*, 541–549.

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. Cell *137*, 571–584.

Ohtsuka, S., and Niwa, H. (2015). The differential activation of intracellular signaling pathways confers the permissiveness of embryonic stem cell derivation from different mouse strains. Development *142*, 431–437.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic Dynamics of Imprinted X Inactivation During Early Mouse Development. Science (80-.). *303*, 644–649.

Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M., and Hamada, H. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell *60*, 461–472.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313–317.

Onishi, K., Tonge, P.D., Nagy, A., and Zandstra, P.W. (2014). Local BMP-SMAD1 signaling increases LIF receptor-dependent STAT3 responsiveness and primed-to-naive mouse pluripotent stem cell conversion frequency. Stem Cell Reports *3*, 156–168.

van Oosten, A.L., Costa, Y., Smith, A., and Silva, J.C.R. (2012). JAK/STAT3 signalling is sufficient and dominant over antagonistic cues for the establishment of naive pluripotency. Nat. Commun. *3*, 817.

Ormsbee Golden, B.D., Wuebben, E.L., and Rizzino, A. (2013). Sox2 Expression Is Regulated by a Negative Feedback Loop in Embryonic Stem Cells That Involves AKT Signaling and FoxO1. PLoS One *8*, e76345.

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat. Biotechnol. *26*, 313–315.

Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P.J., Chambers, I., and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. Development *139*, 2288–2298.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell *6*, 382–395.

Park, S.-H., Park, S.H., Kook, M.-C., Kim, E.-Y., Park, S., and Lim, J.H. (2004). Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells. Ultrastruct. Pathol. *28*, 229–238.

Pasque, V., Radzisheuskaya, A., Gillich, A., Halley-Stott, R.P., Panamarova, M., Zernicka-Goetz, M., Surani, M.A., and Silva, J.C.R. (2012). Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency. J. Cell Sci. *125*, 6094–6104.

Pereira, L., Yi, F., and Merrill, B.J. (2006). Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. Mol. Cell. Biol. *26*, 7479–7491.

Pevny, L.H., and Nicolis, S.K. (2010). Sox2 roles in neural stem cells. Int. J. Biochem. Cell Biol. *42*, 421–424.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A.-K.A.-K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development *135*, 3081–3091.

Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli, S., Tan, K.Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat. Biotechnol. *28*, 848–855.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. Cell *151*, 1617–1632.

Pujadas, E., and Feinberg, A.P. (2012). Regulated noise in the epigenetic landscape of development and disease. Cell *148*, 1123–1131.

Que, J., Okubo, T., Goldenring, J.R., Nam, K.-T., Kurotani, R., Morrisey, E.E., Taranova, O., Pevny, L.H., and Hogan, B.L.M. (2007). Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development *134*, 2521–2531.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raab, S., Klingenstein, M., Liebau, S., and Linta, L. (2014). A Comparative View on Human Somatic Cell Sources for iPSC Generation. Stem Cells Int. *2014*, 768391.

Radzisheuskaya, A. (2014). Mechanisms of induced pluripotency: role of histone variant macroH2A and transcription factor Oct4. University of Cambridge.

Radzisheuskaya, A., and Silva, J.C.R.R. (2014). Do all roads lead to Oct4? The emerging concepts of induced pluripotency. Trends Cell Biol. *24*, 275–284.

Radzisheuskaya, A., Chia Gle, B., dos Santos, R.L., Theunissen, T.W., Castro, L.F., Nichols, J., and Silva, J.C. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nat Cell Biol *15*, 579–590.

Ramos-Mejía, V., Escalante-Alcalde, D., Kunath, T., Ramírez, L., Gertsenstein, M., Nagy, A., and Lomelí, H. (2005). Phenotypic analyses of mouse embryos with ubiquitous expression of Oct4: Effects on mid-hindbrain patterning and gene expression. Dev. Dyn. *232*, 180–190.

Rastan, S., and Robertson, E.J. (1985). X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. Development *90*.

Reyes de Mochel, N.S., Luong, M., Chiang, M., Javier, A.L., Luu, E., Toshihiko, F., MacGregor, G.R., Cinquin, O., and Cho, K.W.Y. (2015). BMP signaling is required for cell cleavage in preimplantation-mouse embryos. Dev. Biol. *397*, 45–55.

Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. Cell Stem Cell *10*, 583–594.

Reznikoff, W.S. (1989). Regulation of Bacterial Gene Expression. In Bacteria in Nature, (Boston, MA: Springer US), pp. 337–358.

Robertson, E., Bradley, A., Kuehn, M., and Evans, M. (1986). Germ-Line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature *323*, 445–448.

Roode, M., Blair, K., Snell, P., Elder, K., Marchant, S., Smith, A., and Nichols, J. (2012). Human hypoblast formation is not dependent on FGF signalling. Dev. Biol. *361*, 358–363.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.-K.K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., Wrana, J.L., Chai, L., et al. (2010). Functional genomics reveals a BMP-Driven mesenchymal-to-Epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell *7*, 64–77.

dos Santos, R.L., Tosti, L., Radzisheuskaya, A., Caballero, I.M., Kaji, K., Hendrich, B., and Silva, J.C.R. (2014). MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. Cell Stem Cell *15*, 102–110.

Sarkar, A., and Hochedlinger, K. (2013). The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell *12*, 15–30.

Schäfer, B.W., Blakely, B.T., Darlington, G.J., and Blau, H.M. (1990). Effect of cell history on response to helix−loop−helix family of myogenic regulators. Nature *344*, 454–458.

Schmitz, J., Weissenbach, M., Haan, S., Heinrich, P.C., and Schaper, F. (2000). SOCS3 exerts its inhibitory function on interleukin-6 signal transduction through the SHP2 recruitment site of gp130. J Biol Chem *275*, 12848–12856.

Schneuwly, S., Klemenz, R., and Gehring, W.J. (1987). Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. Nature *325*, 816–818.

Schöler, H.R., Dressler, G.R., Balling, R., Rohdewohld, H., and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. EMBO J. *9*, 2185–2195.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. Mol. Cell *48*, 849–862.

Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y., and Saitou, M. (2007). Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development *134*, 2627–2638.

Sharova, L. V, Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S.H. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res. *16*, 45–58.

Shimozaki, K., Nakashima, K., Niwa, H., and Taga, T. (2003). Involvement of Oct3/4 in the enhancement of neuronal differentiation of ES cells in neurogenesis-inducing cultures. Development *130*, 2505–2512.

Shipony, Z., Mukamel, Z., Cohen, N.M., Landan, G., Chomsky, E., Zeliger, S.R., Fried, Y.C., Ainbinder, E., Friedman, N., and Tanay, A. (2014). Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. Nature *513*, 115–119.

Shu, J., Wu, C., Wu, Y., Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X., et al. (2013). Induction of Pluripotency in Mouse Somatic Cells with Lineage Specifiers. Cell *153*, 963–975.

Shy, B.R., Wu, C.-I., Khramtsova, G.F., Zhang, J.Y., Olopade, O.I., Goss, K.H., and Merrill, B.J. (2013). Regulation of Tcf7l1 DNA binding and protein stability as principal mechanisms of Wnt/β-catenin signaling. Cell Rep. *4*, 1–9.

Silva, J., and Smith, A. (2008). Capturing Pluripotency. Cell *132*, 532–536.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol. *6*.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737.

Simons, B.D., and Clevers, H. (2011). Strategies for Homeostatic Stem Cell Self-Renewal in Adult Tissues. Cell *145*, 851–862.

Smith, A. (2017a). Formative pluripotency: the executive phase in a developmental continuum. Development *144*, 365–373.

Smith, A.G. (2001). Embryo-Derived Stem Cells: Of Mice and Men. Annu. Rev. Cell Dev. Biol. *17*, 435–462.

Smith, J.J. (2017b). Large-Scale Programmed Genome Rearrangements in Vertebrates. In Somatic Genome Variation in Animals, Plants, and Microorganisms, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), pp. 45–53.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688–690.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339–344.

Smith, Z.D., Sindhu, C., and Meissner, A. (2016). Molecular features of cellular reprogramming and development. Nat. Rev. Mol. Cell Biol. *17*, 139–154.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.

Sousa, E.J., Stuart, H.T., Bates, L.E., Ghorbani, M., Nichols, J., Dietmann, S., and Silva, J.C.R. (2018). Exit from Naive Pluripotency Induces a Transient X Chromosome Inactivation-like State in Males. Cell Stem Cell *22*, 1–10.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., Plath, K., Ruotti, V., Stewart, R., et al. (2009). Role of the murine reprogramming factors in the induction of pluripotency. Cell *136*, 364–377.

Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nat. Cell Biol. *15*, 872–882.

Stewart, M.H., Bossé, M., Chadwick, K., Menendez, P., Bendall, S.C., and Bhatia, M. (2006). Clonal isolation of hESCs reveals heterogeneity within the pluripotent stem cell compartment. Nat. Methods *3*, 807–815.

Stirparo, G.G., Boroviak, T., Guo, G., Nichols, J., Smith, A., and Bertone, P. (2018). Integrated analysis of single-cell embryo data yields a unified transcriptome signature for the human preimplantation epiblast. Development dev.158501.

Sugimoto, M., Kondo, M., Koga, Y., Shiura, H., Ikeda, R., Hirose, M., Ogura, A., Murakami, A., Yoshiki, A., Chuva de Sousa Lopes, S.M., et al. (2015). A simple and robust method for establishing homogeneous mouse epiblast stem cell lines by wnt inhibition. Stem Cell Reports *4*, 744–757.

Sumi, T., Oki, S., Kitajima, K., and Meno, C. (2013). Epiblast Ground State Is Controlled by Canonical Wnt/β-Catenin Signaling in the Postimplantation Mouse Embryo and Epiblast Stem Cells. PLoS One *8*, e63378.

Sutherland, A.E., Speed, T.P., and Calarco, P.G. (1990). Inner cell allocation in the mouse morula: The role of oriented division during fourth cleavage. Dev. Biol. *137*, 13–25.

Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Dev. Biol. *294*, 525–540.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr Biol *11*, 1553–1558.

Tai, C.-I., and Ying, Q.-L. (2013). Gbx2, a LIF/Stat3 target, promotes reprogramming to and retention of the pluripotent ground state. J. Cell Sci. *126*, 1093–1098.

Tai, C.-I., Schulze, E.N., and Ying, Q.-L. (2014). Stat3 signaling regulates embryonic stem cell fate in a dose-dependent manner. Biol. Open *3*, 958–965.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254–1269.

Tam, P.P.L., and Rossant, J. (2003). Mouse embryonic chimeras: tools for studying mammalian development. Development *130*, 6155–6163.

Tang, Y., Luo, Y., Jiang, Z., Ma, Y., Lin, C.-J., Kim, C., Carter, M.G., Amano, T., Park, J., Kish, S., et al. (2012). Jak/Stat3 Signaling Promotes Somatic Cell Reprogramming by Epigenetic Regulation. Stem Cells *30*, 2645–2656.

Teo, A.K.K., Arnold, S.J., Trotter, M.W.B., Brown, S., Ang, L.T., Chng, Z., Robertson, E.J., Dunn, N.R., and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes Dev. *25*, 238–250.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Theunissen, T.W., and Jaenisch, R. (2014). Molecular control of induced pluripotency. Cell Stem Cell *14*.

Theunissen, T.W., Van Oosten, A.L., Castelo-Branco, G., Hall, J., Smith, A., and Silva, J.C.R. (2011). Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. Curr. Biol. *21*, 65–71.

Theunissen, T.W.W., Powell, B.E.E., Wang, H., Mitalipova, M., Faddah, D.A.A., Reddy, J., Fan, Z.P.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency. Cell Stem Cell *15*, 471–487.

Thier, M., Wörsdörfer, P., Lakes, Y.B., Gorris, R., Herms, S., Opitz, T., Seiferling, D., Quandel, T., Hoffmann, P., Nöthen, M.M., et al. (2012). Direct conversion of fibroblasts into stably expandable neural stem cells. Cell Stem Cell *10*, 473–479.

Thomas, K.R., Folger, K.R., and Capecchi, M.R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. Cell *44*, 419–428.

Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L., and Melton, D.W. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell *56*, 313–321.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic Stem Cell Lines Derived from Human Blastocysts. Science (80-.). *282*, 1145–1147.

Thomson, M., Liu, S.J., Zou, L.-N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell *145*, 875–889.

Tokuzawa, Y., Kaiho, E., Maruyama, M., Takahashi, K., Mitsui, K., Maeda, M., Niwa, H., and Yamanaka, S. (2003). Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development. Mol. Cell. Biol. *23*, 2699–2708.

Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M., and Okuda, A. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Res. *30*, 3202–3213.

Toribio, A.L., Alako, B., Amid, C., Cerdeño-Tarrága, A., Clarke, L., Cleland, I., Fairley, S., Gibson, R., Goodgame, N., Ten Hoopen, P., et al. (2017). European Nucleotide Archive in 2016. Nucleic Acids Res. *45*, D32–D36.

Torres, J., and Watt, F.M. (2008). Nanog maintains pluripotency of mouse embryonic stem cells by inhibiting NFkappaB and cooperating with Stat3. Nat Cell Biol *10*, 194–201.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. Development *135*, 909–918.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., et al. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391–395.

Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wymeersch, F., Osorno, R., Economou, C., Karagianni, E., Zhao, S., Lowell, S., et al. (2014). Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. Development *141*, 1209–1221.

Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W.B., Cho, C.H.-H., Martinez, A., Rugg-Gunn, P., et al. (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development *136*, 1339–1349.

Veillard, A.-C., Marks, H., Bernardo, A.S., Jouneau, L., Laloë, D., Boulanger, L., Kaan, A., Brochard, V., Tosolini, M., Pedersen, R., et al. (2014). Stable Methylation at Promoters Distinguishes Epiblast Stem Cells from Embryonic Stem Cells and the In Vivo Epiblasts. Stem Cells Dev.

Voss, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat. Rev. Genet. *15*, 69–81.

Waddington, C.H. (1957). The Strategy of the Genes (Allen & Unwin, London).

Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell *10*, 440–454.

Wegner, M., and Stolt, C.C. (2005). From stem cells to neurons and glia: a Soxist's view of neural development. Trends Neurosci. *28*, 583–588.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat. Genet. *41*, 246–250.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318–324.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wiles, M. V., and Johansson, B.M. (1999). Embryonic Stem Cell Development in a Chemically Defined Medium. Exp. Cell Res. *247*, 241–248.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684–687.

Wilson, N.K., Kent, D.G., Buettner, F., Shehata, M., Macaulay, I.C., Calero-Nieto, F.J., Sánchez Castillo, M., Oedekoven, C.A., Diamanti, E., Schulte, R., et al. (2015). Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell *16*, 712–724.

Wong, F.C.K. (2015). The role of Sox2 in postimplantation epiblast pluripotency. University of Edinburgh.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem. Soc. Trans. *38*, 1027–1032.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat. Cell Biol. *13*, 838–845.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell *153*, 1134–1148.

Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J.A., and Kosik, K.S. (2009). MicroRNA-145 Regulates OCT4, SOX2, and KLF4 and Represses Pluripotency in Human Embryonic Stem Cells. Cell *137*, 647–658.

Yamane, M., Ohtsuka, S., Matsuura, K., Nakamura, A., and Niwa, H. (2018). Overlapping function of klf family targets multiple transcription factors to maintain naïve pluripotency of ES cells. Development dev.162404.

Yang, J., van Oosten, A.L., Theunissen, T.W., Guo, G., Silva, J.C.R.R., and Smith, A. (2010). Stat3 activation is limiting for reprogramming to ground state pluripotency. Cell Stem Cell *7*, 319–328.

Yang, S.-H., Kalkan, T., Morissroe, C., Marks, H., Stunnenberg, H., Smith, A., and Sharrocks, A.D. (2014). Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. Cell Rep. *7*, 1968–1981.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic Acids Res. *44*.

Ye, S., Li, P., Tong, C., and Ying, Q.L. (2013). Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. EMBO J *32*, 2548–2560.

Yeo, J.-C., Jiang, J., Tan, Z.-Y., Yim, G.-R., Ng, J.-H., Göke, J., Kraus, P., Liang, H., Gonzales, K.A.U.A.U., Chong, H.-C., et al. (2014). Klf2 Is an Essential Factor that Sustains Ground State Pluripotency. Cell Stem Cell *14*, 864–872.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development *122*.

Yin, R., Mao, S.Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., et al. (2013). Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. J Am Chem Soc *135*, 10396–10403.

Ying, Q.-L., Nichols, J., Chambers, I., Smith, A., Adelman, C.., Chattopadhyay, S., Bieker, J.., Attisano, L., Wrana, J.., Baonza, A., et al. (2003). BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. Cell *115*, 281–292.

Ying, Q.-L.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Yoshiura, S., Ohtsuka, T., Takenaka, Y., Nagahara, H., Yoshikawa, K., and Kageyama, R. (2007). Ultradian oscillations of Stat, Smad, and Hes1 expression in response to serum. Proc Natl Acad Sci U S A *104*, 11292–11297.

You, J.S., Kelly, T.K., De Carvalho, D.D., Taberlay, P.C., Liang, G., and Jones, P.A. (2011). OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. Proc. Natl. Acad. Sci. U. S. A. *108*, 14497–14502.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940–954.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev. *9*, 2635–2645.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zeineddine, D., Papadimou, E., Chebli, K., Gineste, M., Liu, J., Grey, C., Thurig, S., Behfar, A., Wallace, V.A., Skerjanc, I.S., et al. (2006). Oct-3/4 Dose Dependently Regulates Specification of Embryonic Stem Cells toward a Cardiac Lineage and Early Heart Development. Dev. Cell *11*, 535–546.

Zhang, H., Gayen, S., Xiong, J., Zhou, B., Shanmugam, A.K., Sun, Y., Karatas, H., Liu, L., Rao, R.C., Wang, S., et al. (2016). MLL1 Inhibition Reprograms Epiblast Stem Cells to Naive Pluripotency. Cell Stem Cell *18*, 481–494.

Zhang, M., Leitch, H.G., Tang, W.W.C., Festuccia, N., Hall-Ponsele, E., Nichols, J., Surani, M.A., Smith, A., and Chambers, I. (2018). Esrrb Complementation Rescues Development of Nanog-Null Germ Cells. Cell Rep. *22*, 332–339.

Zhao, S., Nichols, J., Smith, A.G., and Li, M. (2004). SoxB transcription factors specify neuroectodermal lineage choice in ES cells. Mol. Cell. Neurosci. *27*, 332–342.

Zhou, G., Myers, R., Li, Y., Chen, Y., Shen, X., Fenyk-Melody, J., Wu, M., Ventre, J., Doebber, T., Fujii, N., et al. (2001). Role of AMP-activated protein kinase in mechanism of metformin action. J. Clin. Invest. *108*, 1167–1174.

Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., Blau, C.A., Horwitz, M.S., Hockenbery, D., Ware, C., et al. (2012). HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. EMBO J. *31*, 2103–2116.

Zhu, J., Adli, M., Zou, J.Y.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide Chromatin State Transitions Associated with Developmental and Environmental Cues. Cell *152*, 642–654.

Zylicz, J.J., Dietmann, S., Günesdogan, U., Hackett, J.A., Cougot, D., Lee, C., and Surani, M.A. (2015). Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. Elife *4*, e09571.

APPENDIX

Distinct routes converge on a unifying transition logic to establish naïve pluripotency

Hannah T Stuart[1,2], Giuliano G Stirparo[1], Tim Lohoff[1,3], Lawrence E Bates[1,2], Masaki Kinoshita[1], Chee Y Lim[1,4], Elsa J Sousa[1], Katsiaryna Maskalenka[1], Aliaksandra Radzisheuskaya[1,2], Andrew Malcolm[1], Mariana R P Alves[1], Rebecca L Lloyd[1,2], Sonia Nestorowa[1,4], Peter Humphreys[1], William Mansfield[1], Wolf Reik[5], Paul Bertone[1], Jennifer Nichols[1,3], Berthold Göttgens[1,4], José C R Silva[1,2,$]

1. Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, UK
2. Department of Biochemistry, University of Cambridge, UK
3. Department of Physiology, Development and Neuroscience, University of Cambridge, UK
4. Department of Haematology & CIMR, University of Cambridge, UK
5. Babraham Institute & Wellcome Sanger Institute, Cambridge, UK

$ Lead Contact: jcs64@cam.ac.uk

Summary

To decipher how cellular identity is instructed by interplay between transcription factors and signals, we employ defined reprogramming systems in which genetic and signalling parameters can be independently varied and successfully transitioning cells isolated. We show that naïve pluripotency can be induced from EpiSCs along transcriptionally and mechanistically distinct routes. Relative to development, one route moves forward, with productive cells acquiring mesodermal signature prior to naïve pluripotency induction. In contrast, another route overshoots backwards, transcriptionally resembling the earlier embryo and gaining its greater developmental potency. Nevertheless, these distinct trajectories reach the same endpoint, demonstrating surprising flexibility for the establishment of a single identity from a single origin. We reconcile route differences, revealing precise Oct4 expression as a unifying, essential and sufficient feature. We propose that fine-tuned regulation of this 'transition factor' underpins multidimensional access to the naïve identity. This offers a conceptual framework for the understanding of cell identity transitions.

Introduction

Differential use of the same genome generates the spectacular diversity of biological form and function in multicellular animals. A finite number of transcription factors (TFs) and signalling pathways are used and reused in different combinations and contexts in order to generate this array of distinct cellular identities. But how is interplay between external signals and internal TF networks computed by the cell to instruct identity? Are there multiple routes by which a given identity can be established, or must it always follow the same progression of mechanistic steps? These are fundamental questions of wide interest, answers to which will underpin our understanding of multicellular biology.

A cellular identity with a stable gene regulatory network can be thought of as an attractor, occupying a local minimum in an 'energetic' landscape of cell states (Kauffman, 1993; reviewed in Enver et al., 2009). But is an attractor multidimensional, with multiple ways by which it can be approached, or do identity transitions follow a set path through an energetic 'valley'? Empirical evidence supporting theories of cellular identity as a multidimensional attractor was provided in a landmark work by Huang et al., 2005. They showed two transcriptionally distinct routes of promyelocytic HL60 cell differentiation into neutrophils, although they noted some disparity in the resulting neutrophil identities. A limitation for the further understanding of the principles governing cell identity change has been a lack of suitable *in vitro* cell types and of defined, tractable systems to study the transitions occurring between these.

Here we investigate the principles underpinning cell identity transitions. To address this we chose reprogramming to naïve pluripotent stem cells (nPSCs) as a model.

nPSCs have an unbiased potential to make all lineages of the developed organism, including the germ lineage. This fascinating identity arises naturally in the pre-implanta-

tion mammalian epiblast, and can be captured *in vitro* as embryonic stem cells (hereafter referred to as nPSCs), or generated by reprogramming of differentiated cells back into induced nPSCs (inPSCs) (Takahashi & Yamanaka, 2006). Murine naïve pluripotency can be maintained in culture by dual inhibition (2i) of Mek/Erk by PD03 and Gsk3 by Chiron, together with Stat3-agonist LIF (Ying et al., 2008). Core members of the TF network regulating the naïve identity include Oct4, Sox2, Nanog, Esrrb, Klf2, Klf4, Klf5, Stat3 and Tfcp2l1, and multiple inputs have been identified between the 2iLIF signal components and this network (reviewed in Martello & Smith, 2014).

In the post-implantation epiblast, the pluripotent cells have progressed to the primed state. This distinct identity exhibits markedly different transcriptional, epigenetic and metabolic profiles, and no longer gives rise to the germ lineage (reviewed in Morgani et al., 2017). These cells can be captured in culture as epiblast stem cells (EpiSCs) and require FGF stimulation rather than inhibition of Mek/Erk signalling, together with the addition of ActivinA (FA) (Brons et al., 2007; Tesar et al., 2007).

Reprogramming of EpiSCs to inPSCs provides several advantages as a model system to study cell identity transitions. The destination naïve identity is extremely well defined in terms of its molecular signature, while functional assays such as clonogenic expansion and chimeric contribution leave no doubt as to whether the identity in question has indeed been generated. Reprogramming of EpiSCs requires only one driving naïve factor, combined with defined modulation of the signalling environment (Guo et al., 2009; van Oosten et al., 2012). This is in stark contrast to somatic cell reprogramming, which requires multiple genetic and signal variables to be introduced simultaneously in order to achieve reprogramming, prohibiting causal ascription of changes to individual inputs (reviewed in Smith, Sindhu & Meissner, 2016). Furthermore, rapid naïve gene expression responses follow transgene induction in EpiSCs, even while maintaining EpiSC FA culture conditions (Stuart et al., 2014). Thus, in this system we can disentangle the contributions of TFs and the environment to identity induction mechanisms and kinetics.

By use of individual, inducible factors coupled with independent manipulation of environmental parameters, we interrogated how the naïve pluripotent identity is instructed by interplay between TFs and signals. As a result we defined principles and new mechanisms governing naïve pluripotency establishment, which were also applicable to other contexts including embryonic development and somatic cell reprogramming. Importantly, we provide explicit evidence for cellular identity as a multidimensional attractor state, with mechanistically as well as transcriptionally distinct pathways to transit between the same start and end identities.

Results

Reprogramming initiation is driver-dependent

In order to causally ascribe independent genetic and signal variables to reprogramming events, use of single drivers is necessary. Therefore, we tested reprogramming efficacy of individual naïve factors in embryo-derived *Rex1*$^{+/dGFP.IRES.bsd}$ (*Rex1::*dGFP) EpiSCs (Figure 1A–C). Dox-inducible (i) transgenes were used for Esrrb, Klf2, Klf4, Nanog, Tfcp2l1 and Klf5. Stat3 activation by phosphorylation (iPStat3) was elicited by GCSF-stimulation of the GY118F receptor transgene (Burdon et al., 1999), since LIF signal transduction of EpiSCs is limited (Yang et al., 2010). iEsrrb, iPStat3 and iKlf2 were the three most efficient single drivers in 2iLIF (Fig 1C). Interestingly, each inputs to the naïve network along a different regulatory axis (Fig 1D), that of Chiron, LIF and PD03 respectively (Martello et al., 2012; Niwa et al., 1998; Yeo et al., 2014). inPSCs established from these lines were transcriptionally indistinguishable (Fig 1E, S1A) and were chimera and germline competent (Fig 1F), demonstrating molecular and functional equivalency. Therefore, we took iEsrrb, iPStat3 and iKlf2 as a complementary model set of single reprogramming drivers for mechanistic study.

We analysed the initial transcriptional response to each driver from 1–48h (Fig 1G). In 2iLIF, naïve gene upregulation by iPStat3 was moderate, by iEsrrb was substantially faster and stronger, while iKlf2 surprisingly did not upregulate naïve genes and even silenced Sox2 (Fig 1G). These differing kinetics

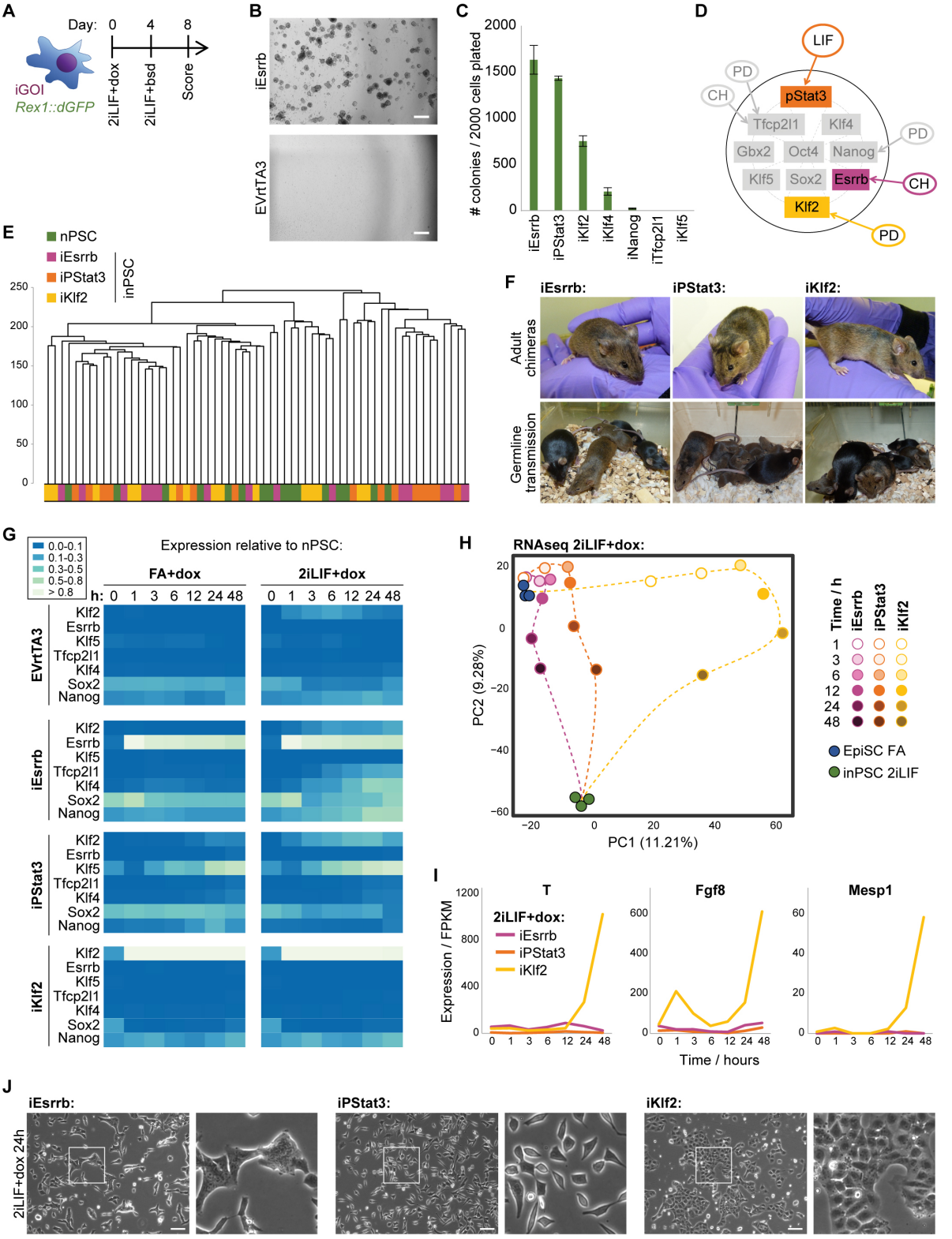


Figure 1: Reprogramming initiation is driver-dependent

(A) Reprogramming protocol for *Rex1*$^{+/dGFP.IRES.bsd}$ (*Rex1*::dGFP) EpiSCs with individual, inducible driver genes of interest (iGOI). Bsd=blasticidin. **(B)** Phase images of iEsrrb and EmptyVector+rtTA3 (EVrtTA3, negative control) wells on day 8. Scale: 500µm. **(C)** Mean number of inPSC colonies on day 8 ±SD (n=3) per 2000 cells plated. **(D)** Inputs of Esrrb, PStat3 and Klf2 to the naïve network. Signals: PD=PD03, CH=Chiron. **(E)** Unsupervised hierarchical cluster of scRNAseq, computed with Ward.D2 agglomeration method and Euclidean distances for all expressed genes (FPKM>0). **(F)** Blastocyst injection of inPSCs (agouti) yielded high-contribution adult chimeras, capable of germline transmission (agouti pups). **(G)** Heatmap of mean gene expressions from 0–48h, measured by RT-qPCR relative to Gapdh then normalised to nPSCs. **(H)** PCA based on variable genes ($\log_2$ FPKM>1, CV^2>0.5 calculated for each driver then merged to a single list). **(I)** Expression of mesodermal markers following reprogramming induction. **(J)** Images and zooms 24h after induction. Scale: 100µm. **See also Figure S1.**

are further reflected by the rates of *Rex1::*dGFP upregulation and the rates of formation of transgene-independent inPSCs from day 2 onwards (Fig S1B–C), yet are not attributable to differences in the kinetics or levels of transgene induction (Fig S1D–E).

We compared transgene induction in 2iLIF or FA conditions from 1–48h, to assess the contribution of and interplay between genetic and signal variables. For iPStat3 and iKlf2, responses were similar in EpiSC FA and 2iLIF conditions (Fig 1G). However, for iEsrrb the response was highly condition-dependent, suggesting that Esrrb and 2iLIF work in synergy to rapidly induce naïve genes. Indeed, iEsrrb and 2iLIF components interact to elicit a transcriptional trajectory distinct from that of iEsrrb in EpiSC FA conditions (Fig S1F). The signalling environment did not play a strong role in the early transcriptional behaviour of iKlf2, with more similarity between timepoints than conditions (Fig S1F).

Considering that Klf2 is a potent reprogramming driver (Fig 1C), its initial lack of naïve gene induction presented a fascinating conundrum. Principle component analysis (PCA) showed a remarkable transcriptional divergence following Klf2 induction, corroborated by *k*-means cluster analysis (Fig 1H, S1G). We asked which genes could cause such a transcriptional divergence, and found robust upregulation of mesodermal markers in a Klf2-specific manner (Fig 1I). This indicates initial instigation of a different program downstream of Klf2, rather than simply a delayed naïve induction kinetic.

Together, expression analyses revealed that the pattern and kinetics of naïve network reactivation were driver-dependent, and that signal contribution was modulated by the driver. Morphological changes during reprogramming initiation were also driver-specific (Fig 1J). Nonetheless, these divergent processes ultimately re-converged on the same naïve pluripotent destination identity (Fig 1E–F, S1A).

Single-cell RNAseq defines distinct productive trajectories

Since reprogramming to naïve pluripotency is heterogeneous and asynchronous, cells undergoing the change of interest must be resolved from the average in order to study transition mechanisms (Fig 2A) (reviewed in Buganim et al., 2013). Therefore, we tested isolation of productively transitioning intermediates based on activation of *Rex1*::dGFP reporter. *Rex1* is silent in EpiSCs, increases incrementally during reprogramming (Stuart et al., 2014), and is extensively characterised as a sensitive proxy of naïve network strength (Kalkan et al., 2017). When replated in 2iLIF+dox/GCSF, we found that emergent dGFP+ reprogramming intermediates were destined to form naïve colonies with efficiency comparable to nPSCs (Fig 2B).

We performed single-cell (sc)RNAseq at 12 and 24h (all cells), on dGFP-/low/+ at 72h, and on dGFP+ at 48, 72 and 96h (Fig 2C). With the former we capture early differences and trajectory overviews, while the 48–96h dGFP+ resolves cells undergoing productive progression to naïve pluripotency. PCA revealed that, for iEsrrb and iPStat3, the start EpiSCs and end inPSCs represent the extremes of identity along PC1. In contrast, in the first 12–24h, iKlf2 shows a marked diversion, away from both EpiSC and inPSC along PC1 (Fig 2D) and corroborated by unsupervised hierarchical clustering (Fig S2A). To investigate the molecular features of this early diversion, we performed Gene Ontology (GO) analysis (Table S1). There was significant GO enrichment for processes involved in cell motility and development, consistent with initial diversion of iKlf2 in a mesodermal direction.

To further investigate the trajectory distinctions, we performed differential gene expression (DE) analysis. We compared each sample with the start EpiSCs, to see how the expression signatures changed over time for each driver, and, by using a common reference, to assess similarities vs differences between drivers. We plotted Venn diagrams to find the numbers of DE genes that are unique to or shared between drivers at each timepoint (Fig S2B) and summarise these DE numbers in Fig 2E. These clearly show that drivers initially diverge, in particular with iKlf2 exhibiting 2985 unique DE genes at 12h. Over time, the drivers then reconverge, indicated by the increasing proportion of shared DE genes. At 72h, there is positive correlation between dGFP level and the proportion of shared DE genes, consistent with approach of distinct trajectories to the common destination identity.

The initial divergence of iKlf2 cannot simply be attributed to an unproductive offshoot. The iKlf2 72h dGFP- cells cluster back

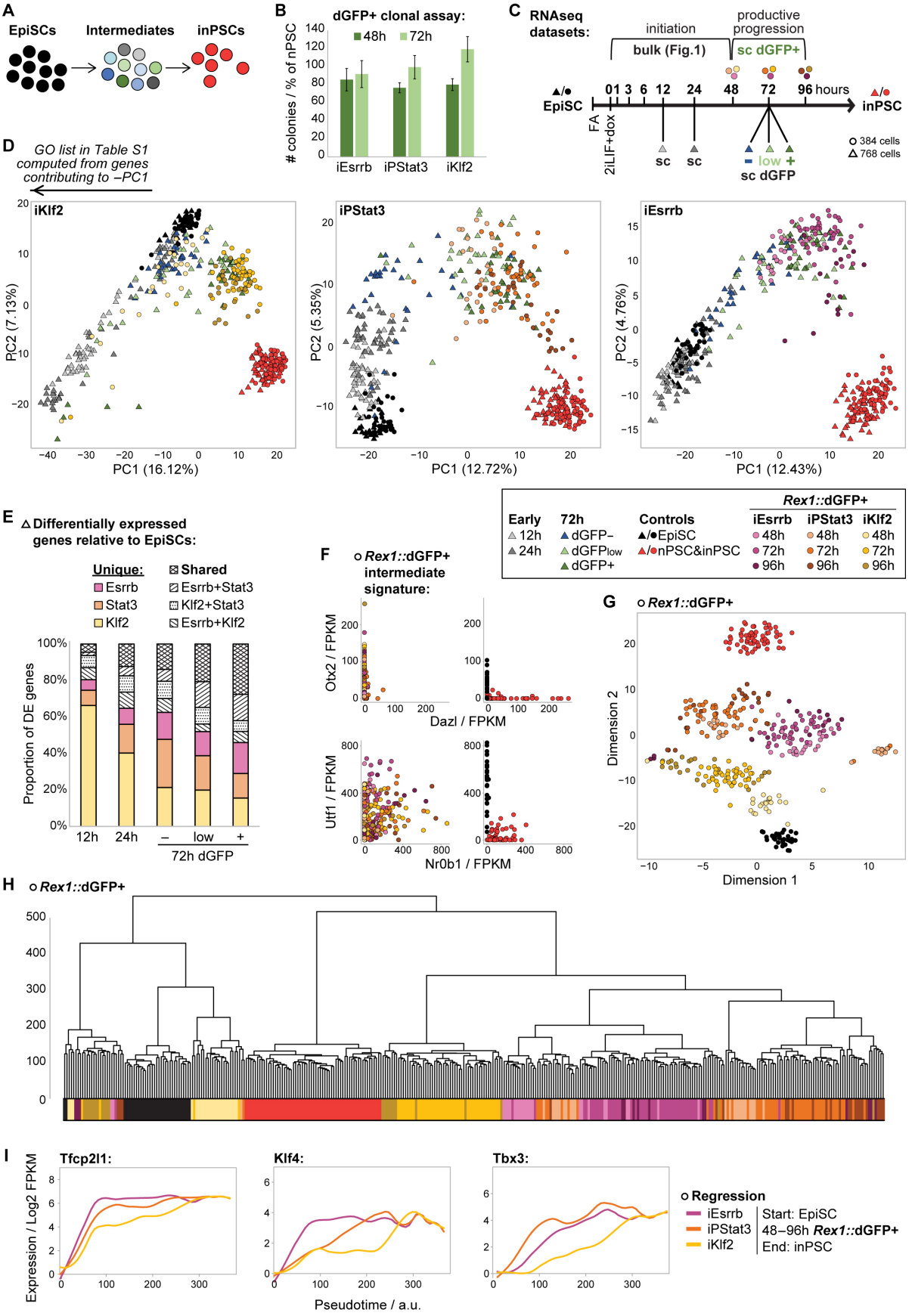


Figure 2: Single-cell RNAseq defines distinct productive trajectories

(A) A method is required to isolate productive intermediates for mechanistic study. **(B)** *Rex1::*dGFP+ cells were isolated by FACS at 48/72h and plated for clonal assay: reprogramming intermediates were plated in 2iLIF+dox, established *Rex1::*dGFP nPSCs in 2iLIF. Dox was withdrawn and blasticidin applied on day 6. Mean numbers of

inPSC colonies ±SD (n=3) scored on day 9 are indicated as % of nPSC colonies for each experiment. **(C)** Schematic summarising RNAseq datasets. sc: single-cell. **(D)** PCA based on variable genes (log_2 FPKM>1, CV2>0.5). **(E)** Numbers of unique and shared differentially expressed (DE) genes for each driver compared to EpiSCs. **(F)** Expression scatter plots of EpiSC markers (Otx2, Utf1) and naïve markers (Dazl, Nr0b1). **(G)** t-SNE plot showing relationships between single cell transcriptomes. **(H)** Unsupervised hierarchical cluster computed with Ward.D2 agglomeration method and Euclidean distances. **(I)** LOESS regression fit lines summarise expression kinetics, computed from scatter plots of log_2 FPKM vs pseudotime for single cells (Fig S2C–D). **See also Figure S2.**

near the EpiSCs, not at the end of a different trajectory (Fig 2D). By live-imaging we confirmed that iKlf2 cells at 12/24h are not undergoing cell death (Video S1). It logically follows that iKlf2 cells start on a divergent trajectory, prior to acquisition of naïve pluripotency.

To connect early trajectory divergence with subsequent acquisition of naïve pluripotency, we analysed the 48–96h dGFP+ cells in more detail. Intermediate identity was confirmed by naïve vs EpiSC marker expression profiles (Fig 2F). Sample relationships determined by t-distributed Stochastic Neighbour Embedding (t-SNE) dimensionality reduction and by hierarchical clustering revealed that dGFP+ sorted intermediates arranged by driver rather than timepoint (Fig 2G–H). This demonstrates that reprogramming routes are transcriptionally distinct throughout the productive transitions, not only during initiation. Again, the iKlf2 route was transcriptionally more different from those of iPStat3 and iEsrrb (Fig 1H, 2H, S2A).

We then examined the kinetics of naïve network activation in single dGFP+ cells. In order to deconvolute the asynchronous nature of reprogramming, we ordered the cells by fraction of similarity to origin EpiSCs and destination inPSCs to assign pseudotime coordinates (Fig S2C). iEsrrb exhibited the fastest kinetics of naïve network orchestration for the majority of naïve genes, while iKlf2 was slowest (Fig 2I, S2D). This is in agreement with the different kinetics observed in bulk analyses from 0–48h (Fig 1G), now extended to 48–96h and within dGFP+ single cells.

iKlf2 reprogramming proceeds via a mesoderm-like state

For iKlf2, the upregulation of mesodermal markers observed during bulk initiation persisted in productive *Rex1::*dGFP+ single-cells (Fig 1I, 3A, S3A). This suggests that transient activation of mesodermal markers was not due to differentiation of a population of unproductive cells, but was a transcriptional response occurring during productive establishment of naïve pluripotency when driven by Klf2. T is specifically expressed in and essential for nascent mesoderm formation. To determine the proportion of iKlf2 intermediates expressing T on the protein level, we performed and quantified immunofluorescence staining following iKlf2 induction (Fig 3B). By 48h, we observe robust expression of T protein in 60% of iKlf2 cells, indicating that these are a major population.

To trace the outcome of these T+ intermediates through the reprogramming process, we generated *T/Rex1* double-reporter EpiSCs (Fig 3C). Into *T::*GFP reporter nPSCs ($T^{+/GFP}$, Fehling et al., 2003), we knocked mKO2 (Orange) into the *Rex1* locus (Fig S3B). We obtained *T/Rex1* double-reporter EpiSCs (TGRO) by differentiation for 10 passages in FA, into which we transfected iKlf2 reprogramming driver. We confirmed that these EpiSCs upregulate T in response to iKlf2 induction, and verified that T and GFP expressions are in agreement (Fig 3D, S3C).

By live-imaging, we traced the activity of *T* and *Rex1* during iKlf2-driven reprogramming of these double-reporter EpiSCs (Fig 3E, Video S2). We observe that T+ colonies emerge around day 2. Strikingly, these T+ colonies then convert into Rex1+ colonies around day 4. The largely sequential nature of *T* then *Rex1* reporter activation is consistent with the low % of T+ cells captured by scRNAseq of Rex1+ intermediates (Fig 3A). Together, this provides direct evidence that productive iKlf2 reprogramming proceeds via a T+ state on the protein level, demonstrating diversion towards mesoderm prior to acquisition of naïve pluripotency.

iPStat3 reprogramming proceeds via an early ICM-like state

To place the reprogramming trajectories in the context of development, we compared scRNAseq of productive *Rex1::*dGFP+ intermediates with E2.5–E6.5 embryos (Deng et al., 2014; Mohammed et al., 2017). Single-cell transcriptome analyses revealed that iPStat3 reprogramming intermediates transiently acquired significant similarity to the early

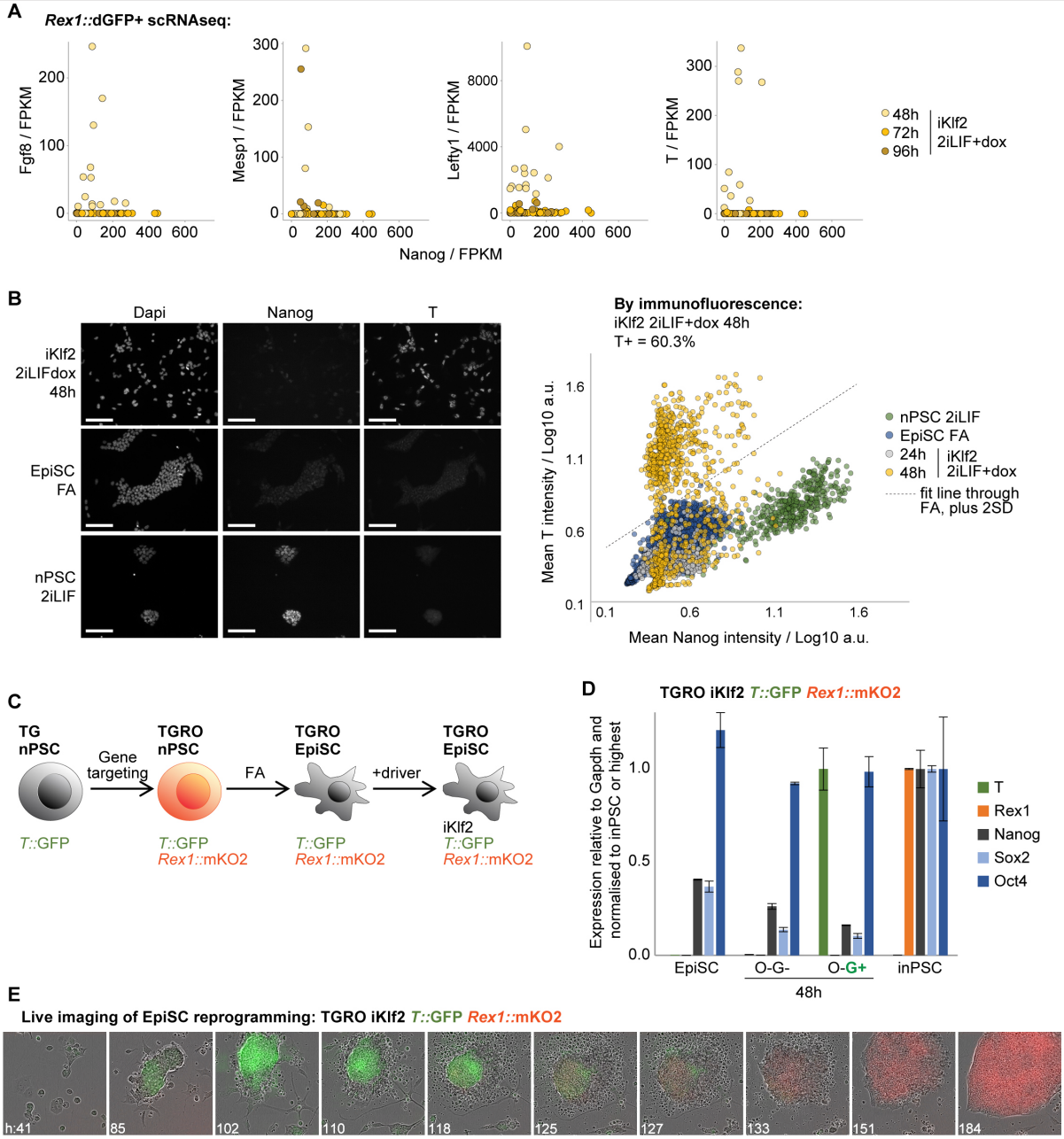


Figure 3: iKlf2 reprogramming proceeds via a mesoderm-like state

(A) Expression scatter plots of mesodermal markers vs Nanog. **(B)** Immunofluorescence staining for T and Nanog was quantified 24/48h after iKlf2 induction in the original *Rex1::*dGFP EpiSCs, on a total of 3675 cells. To determine % of T+ cells a stringent threshold was calculated: mean of EpiSC values +2SD (standard deviations). Scale: 100µm. **(C)** Strategy to generate *T/Rex1* double-reporter (TGRO) iKlf2 EpiSCs. **(D)** RT-qPCR analyses following reprogramming induction of TGRO iKlf2 EpiSCs. *T::*GFP+ (G+) and *T::*GFP- (G-) populations were both *Rex1::*mKO2- (O-) at 48h. Mean expressions are displayed ±SD (n=3). **(E)** Live imaging of TGRO iKlf2 EpiSC reprogramming. On day 4, iKlf2 induction was withdrawn and blasticidin added to select for inPSCs with active *Rex1* reporter. Merge snapshots are shown from Video S2. **See also Figure S3.**

inner cell mass (ICM) (Fig 4A–B) and exhibited a Nanog+Gata6+ double-positive signature (Fig 4C, S4A). Nanog+Gata6+ co-expression is a hallmark of the early ICM (Plusa et al., 2008), prompting the hypothesis that iPStat3-driven reprogramming goes further back to an early ICM-like state, then forwards into the consolidated naïve identity. Indeed, the temporal sequence of naïve gene activation in iPStat3 intermediates emulates that of the embryo (Fig S4B).

In order to functionally test the properties of Gata6+ iPStat3 reprogramming intermediates, we generated *Gata6* reporter EpiSCs by differentiation from *Gata6*$^{+/H2BVenus}$ nPSCs (Freyer et al., 2015). We confirmed that resultant EpiSCs upregulate Gata6 in response

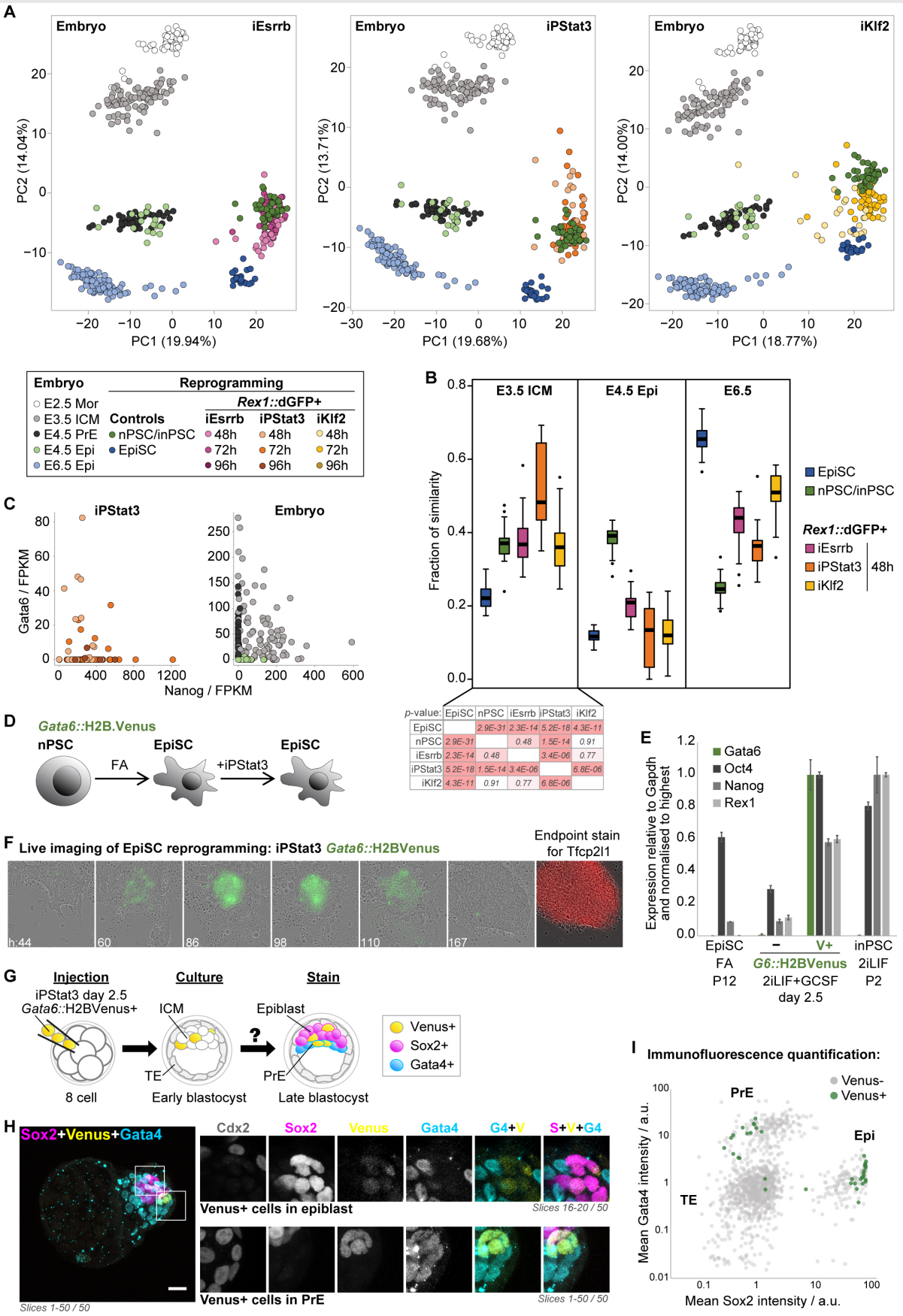


Figure 4: iPStat3 reprogramming proceeds via an early ICM-like state

(A) PCA based on variable genes (log_2 FPKM>1, CV^2>0.5) for reprogramming intermediates and embryo single cells. Mor: compacted morula. ICM: inner cell mass. Epi: epiblast. PrE: primitive endoderm. PC1 separates *in vivo* vs *in vitro* datasets; PC2 portrays developmental progression. **(B)** Fraction of similarly to signature embryo datasets was computed by quadratic programming for each *in vitro* single cell and is presented as box-and-whisker plots.

(C) Scatter plots of Gata6 vs Nanog for iPStat3 reprogramming, E3.5 and E4.5 embryos. **(D)** Strategy to generate *Gata6*-reporter iPStat3 EpiSCs. **(E)** RT-qPCR analyses following reprogramming induction of *Gata6::*H2BVenus iPStat3 EpiSCs. Mean expression is displayed ±SD (n=3). **(F)** Live imaging of *Gata6::*H2BVenus iPStat3 EpiSC reprogramming. On day 4, iPStat3 induction was withdrawn. Merge snapshots are shown from Video S3. Endpoint staining identified inPSC colonies. **(G)** *Gata6*::H2BVenus+ iPStat3 day 2.5 reprogramming intermediates were injected into 8-cell stage embryos then traced in resultant late blastocyst chimeras. **(H)** Maximum intensity Z-projections for a stained chimeric blastocyst. Scale: 20μm. Zooms are shown of indicated regions and slices. Upper: contribution of injected cells to Sox2+Gata4- epiblast is apparent. Although Gata6 is no longer expressed in E4.5 epiblast, Venus has a long half-life allowing us to trace the contribution 2 days after injection of positive cells. Lower: region with contribution of Venus+ cells to Sox2-Gata4+ PrE. Since Gata6 is still expressed in E4.5 PrE, contributing cells actively express Venus. **(I)** Quantification of Gata4 vs Sox2 staining in 7 embryos. **See also Figure S4.**

to iPStat3 induction, that Gata6 and Venus expressions are in agreement, and that Nanog+Gata6+ double-positive cells are present on the protein level (Fig 4D–E, S4C). By live-imaging, we observed that Gata6+ cells emerge at day 2–3 (Fig 4F, Video S3). These subsequently gave rise to inPSCs by the endpoint, providing direct evidence that Gata6+ iPStat3 reprogramming intermediates are productive.

The defining functional property of the early ICM is the potential to generate primitive endoderm (PrE) as well as pluripotent epiblast. To test whether they acquire this greater potency, we injected Gata6+ iPStat3 reprogramming intermediates into 8-cell stage embryos, then cultured these to the late blastocyst stage by which time PrE and epiblast lineages are fully segregated. Chimeric embryos were fixed and analysed for contribution of injected cells to the epiblast (Sox2+), PrE (Gata4+) and trophectoderm (Cdx2+) (Fig 4G). Remarkably, the Gata6+ population contributed to both epiblast and PrE, consistent with a gain of potency equivalent to that of the early ICM (Fig 4H–I). The Gata6+ intermediates were Sox2+Gata4– prior to injection, as the early ICM would be, then could subsequently become either Sox2+Gata4– epiblast or Sox2–Gata4+ PrE in the embryo (Fig S4D). As expected, established inPSCs contributed only to epiblast, and EpiSCs did not contribute at all (data not shown).

In sum, the iPStat3 reprogramming population transiently gains resemblance to the early ICM both in terms of its molecular signature and its developmental potency.

Routes have distinct genetic and signal requirements

To test whether the divergent transcriptional trajectories are indicative of mechanistic differences, we assessed the routes' genetic and signal requirements. Putative downstream genetic mediators were identified by examining expression of known reprogramming drivers 24h after induction of iEsrrb, iKlf2 or iPStat3 (Fig 5A). Endogenous Esrrb was not strongly upregulated by either iKlf2 or iPStat3 by 24h, and correspondingly its knockdown (KD) did not prevent reprogramming (Fig S5A–B). In contrast, endogenous Klf2 reached 50% and 20% of nPSC-level in iPStat3 and iEsrrb respectively. Given this early response, we tested whether Klf2 is a mediator of iPStat3- or iEsrrb-driven reprogramming. We found that transient Klf2 KD abolished reprogramming driven by iPStat3 but not iEsrrb (Fig 5B, S5A–B). This implicates Klf2 as a critical mediator of reprogramming initiation by iPStat3. Klf2 is not considered a PStat3 target in nPSCs, implying different network topologies during establishment versus maintenance of naïve pluripotency. Curiously, iPStat3 sensitivity to Klf2 KD was context-dependent, partially alleviated in the absence of PD03 (Fig S5C).

To assess differences between routes in terms of exogenous signal requirements, we challenged the first 4 days of reprogramming with different 2iLIF signal permutations (Fig 5C). iPStat3 yielded inPSCs in the absence of both PD03 and Chiron, but together PD03 and Chiron synergistically boosted the efficiency. However, the impact of PD03 and Chiron was driver-dependent: Chiron was essential for iKlf2-driven reprogramming, with no benefit from additional supplementation with PD03. Functional redundancy between Klf2 and PD03 has been previously noted (Yeo et al., 2014), and the inability of iKlf2 to drive reprogramming without direction from an exogenous signal is in agreement with the observation that iKlf2 does not directly induce naïve gene expression (Fig 1G). Unlike iKlf2, reprogramming driven by iEsrrb was highly LIF-de-

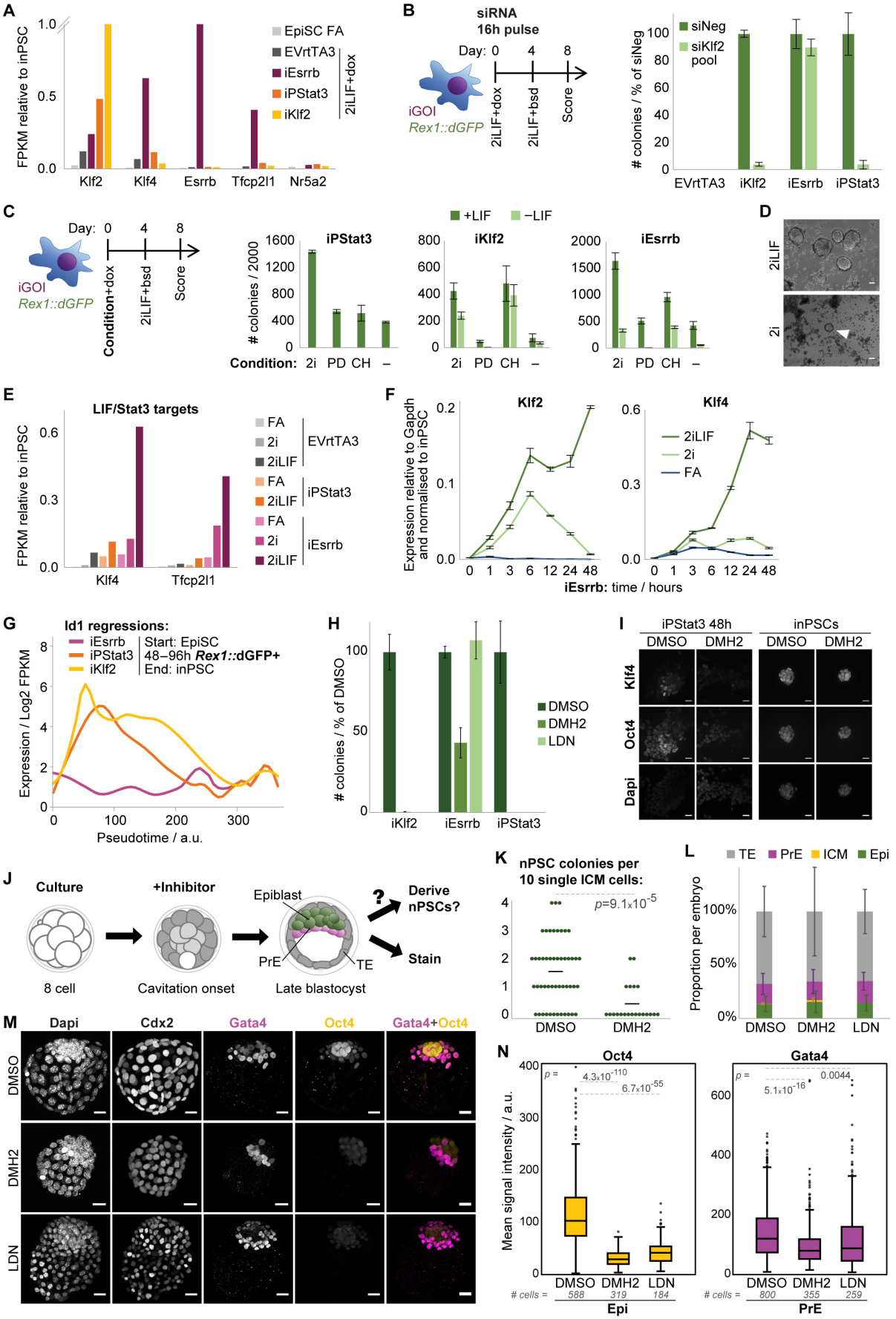


Figure 5: Routes have distinct genetic and signal requirements

(A) Gene expression after 24h relative to inPSCs. *y*-axis iEsrrb: Esrrb=3.32. iKlf2: Klf2=8.30. **(B)** KD was performed at reprogramming onset with a single pulse of siRNA. Mean inPSC colonies scored on day 8 are presented ±SD (n=3). **(C)** Reprogramming was induced in different conditions from day 0–4, then selected in 2iLIF+blasticidin.

inPSC colonies scored on day 8 are presented as mean ±SD (n=3). PD=PD03; CH=Chiron; 2i=PD+CH. **(D)** Phase images of iEsrrb on day 8 in 2iLIF+blasticidin, after reprogramming from day 0–4 in 2iLIF+dox or 2i+dox as indicated. White arrow indicates an inPSC colony. **(E)** Expression of LIF/Stat3 target genes 24h after driver induction in the indicated conditions. **(F)** Timecourse RT-qPCR analyses of iEsrrb EpiSCs in the indicated conditions +dox. Mean expressions are displayed ±SD (n=3). **(G)** LOESS regression fit lines summarise Id1 kinetics during reprogramming, computed from $\log_2$ FPKM vs pseudotime for single cells (Fig S2C). **(H)** 3μM DMH2, 0.6μM LDN or DMSO were applied to reprogramming in 2iLIF+dox/GCSF from day 0–4, then inPSCs were selected in 2iLIF+blasticidin. inPSC colonies scored on day 8 are presented as mean ±SD (n=3). **(I)** Immunofluorescent staining after 48h of inhibitor treatment, for iPStat3 reprogramming in 2iLIF+GCSF or for previously established iPStat3 inPSCs in 2iLIF. Scale: 20μm. **(J)** Schematic summarising BMP inhibitor treatment of pre-implantation embryos. **(K)** Quantitative nPSC-derivation following embryo treatment with DMSO or 3μM DMH2. nPSC colonies were scored per 10 single ICM cells. Black line=mean. DMSO n=7, DMH2 n=8 embryos. **(L–N)** Late blastocysts were stained for Cdx2, Gata4 and Oct4, following treatment with DMSO, 3μM DMH2 or 0.3μM LDN. **(L)** Mean cell number per lineage ± SD, presented as a proportion of the total cells per embryo. DMSO n=23, DMH2 n=18, LDN n=7 embryos. **(M)** Representative maximum intensity Z-projections. Scale: 20μm. **(N)** Quantification of immunofluorescent signal for Oct4 in Epi nuclei and Gata4 in PrE nuclei, presented as box-and-whisker plots. DMSO n=23, DMH2 n=18, LDN n=7 embryos. **See also Figure S5.**

pendent (Fig 5C–D). iEsrrb induction in LIF led to greater upregulation of canonical PStat3 targets than induction of iPStat3 itself (Fig 5E). This was not due to elevation of PStat3 protein by Esrrb (Fig S5D) and thus demonstrates downstream synergy between Esrrb and PStat3. To identify when the synergy between Esrrb and PStat3 became effective, we performed timecourse expression analyses. A turning point occurred 6h after Esrrb induction: from 0–6h, Klf2 was upregulated similarly in 2i±LIF for both iEsrrb and negative control EpiSCs; after 6h, Klf2 expression continued to increase in iEsrrb+2iLIF, but collapsed in iEsrrb+2i and all control conditions (Fig 5F, S5E). Klf4 upregulation also launched in earnest after 6h with iEsrrb+2iLIF.

In light of the above observations that signal requirement and interpretation is driver-dependent, we interrogated *Rex1::*dGFP+ 48–96h scRNAseq data for evidence of other signalling differences between iKlf2, iEsrrb and iPStat3 productive intermediates. BMP signalling pathway target Id1 is upregulated in iKlf2 and iPStat3, but not iEsrrb (Fig 5G). Id1 upregulation is intermediate-specific, with negligible expression in starting EpiSCs or destination inPSCs. BMP signalling is the key pluripotency-sustaining component in the serum of classical nPSC cultures (Ying et al., 2003), is important for MET in serum-based fibroblast reprogramming (Samavarchi-Tehrani et al., 2010), but is not thought to be active in 2iLIF-cultured nPSCs (Boroviak et al., 2014). We assessed BMP pathway status by PSmad1/5 immunofluorescence during EpiSC reprogramming in 2iLIF, finding positive staining for iKlf2 and iPStat3 but not iEsrrb (Fig S5F). Therefore, BMP signalling is activated in a route-specific manner.

To test whether auto/paracrine BMP signalling is required during EpiSC reprogramming, we applied BMP inhibitor from day 0–4. DMH2 is a specific and well-characterised BMP receptor inhibitor (Fig S5G) (Hao et al., 2010), and we also verified key findings with a different inhibitor LDN193189 (LDN) (Cuny et al., 2008). BMP inhibition abolished iKlf2- and iPStat3-driven reprogramming in 2iLIF, but inPSC colonies still formed for iEsrrb (Fig 5H). Therefore, BMP inhibition blocked reprogramming only in those lines exhibiting evidence of active BMP signalling in their intermediates. This was specific to the transition, being dispensable for maintenance of resultant inPSCs in 2iLIF (Fig 5I, S5H).

Together, these results demonstrate that iKlf2, iPStat3 and iEsrrb drive reprogramming by mechanistically distinct routes in terms of their genetic and signal requirements, and their differential modulation of exogenous and endogenous signal transduction.

BMP signalling is required for naïve pluripotency establishment *in vivo*

Having identified BMP signalling requirement in two routes of reprogramming, and given that iPStat3 reprogramming intermediates transiently acquired similarity to the early ICM, we explored whether endogenous BMP signalling also plays a role in naïve pluripotency establishment *in vivo.* The BMP signalling pathway is active in pre-implantation mouse embryos from the 4-cell stage onwards, including in the ICM (Graham et al., 2014; Reyes de Mochel et al., 2015), so in-

volvement in epiblast specification is plausible. We applied BMP inhibitor to the late morula, cultured the embryos to the late blastocyst stage, then analysed the impact on each lineage and performed quantitative nPSC derivation (Fig 5J). Per cell, we observed a 4-fold reduction in nPSC derivation efficiency from embryos that had previously been treated with BMP inhibitor (p=9.1x10^{-5}), demonstrating that BMP inhibition had disrupted pluripotency establishment in the embryo (Fig 5K, S5I).

By analysis of immunofluorescence staining, we counted the number of cells in the Epi, PrE and trophectoderm (TE) lineages, and quantified the intensity of lineage marker expression (Fig 5L–N, S5J–L). The proportions of cells assigned to each lineage were unaffected by BMP inhibition (Fig 5L). PrE and TE exhibited either mildly reduced (DMH2) or unaffected (LDN) lineage marker expression, whereas Oct4 expression in the Epi lineage was dramatically reduced by both inhibitors (p=4.3x10^{-110} for DMH2; p=6.7x10^{-55} for LDN) (Fig 5M–N, S5K). We also performed Nanog staining on a subset of embryos, and observed significant reduction in the Epi for both inhibitors (Fig S5L).

In sum, we found that BMP inhibition gave a specific impact on naïve pluripotency establishment in the embryo, dramatically reducing Epi marker expression and the functional ability to yield nPSCs, despite a normal proportion of cells being allocated to the Epi compartment. Identification of this novel role for the BMP pathway *in vivo* highlights the power of our defined reprogramming systems to uncover principles of identity specification.

Defined Oct4 level is a common feature of all routes

The aforementioned differences in transcriptional trajectories, signal and genetic requirements demonstrate that iKlf2, iPStat3 and iEsrrb instruct reprogramming by distinct mechanisms (Fig 6A). Given that the starting and destination cellular identities are the same in all three cases (Fig 1E–F), the extent of the route differences was surprising. Therefore, we asked whether there was a common feature that could reconcile the disparate transition logics.

From 48–96h in *Rex1::*dGFP+ single cells, we found that Oct4 is expressed at endogenous pluripotent level, irrespective of the driver (Fig 6B). Maintenance of Oct4 throughout the transitions is not to be taken for granted. Although Oct4 is expressed at similar levels in EpiSCs and inPSCs (Fig S6A), this expression is supported by different transcriptional networks and is driven from different enhancer elements (Tesar et al., 2007; Yeom et al., 1996). Indeed, signal switch of control EpiSCs to 2iLIF triggered Oct4 downregulation (Fig 6C). In contrast, Oct4 was unperturbed in nPSCs upon switch from serum+LIF to 2iLIF (Fig 6D), indicating that 2i itself did not suppress Oct4 in a context where cellular identity was constant. Timecourse RT-qPCR analyses showed that Oct4 was expressed at or above pluripotent stem cell (PSC) level in the dGFP+ reprogramming subpopulation from 48h onwards but not always in the dGFP– (Fig S6B). Together, this suggests that signal-mediated collapse of the primed network prior to naïve network construction leads to Oct4 expression loss, creating a 'vulnerable window' between different self-renewing Oct4-supporting configurations. Since 2iLIF triggered Oct4 collapse in control EpiSCs (Fig 6C), we reason that the observed maintenance of Oct4 in 2iLIF during productive reprogramming is an active process coordinated by the driving transgene (Fig 6B).

To evaluate the relationship between Oct4 level and productive reprogramming, we subdivided intermediate populations based on a finer gradient of *Rex1::*dGFP, measured Oct4 expression and replated for clonogenicity assay in 2iLIF. Average Oct4 expression positively correlated with the subsequent reprogramming efficiency of a given subpopulation (Fig 6E, S6C). To test whether this Oct4 maintenance is required for reprogramming, we performed transient Oct4 KD by a single pulse of siRNA treatment at reprogramming onset. inPSC formation was abolished (Fig 6F, S6D).

Fixed Oct4 expression is sufficient for naïve instruction in minimal conditions

Having demonstrated that Oct4 maintenance is observed in and required for productive reprogramming, next we asked whether Oct4 maintenance is sufficient. We generated *Oct4* null EpiSCs that constitutively express ectopic Oct4 at endogenous PSC level (FixedOct4) (Fig 6G–I, S6E), according to methodology described by Radzisheuskaya et al.,

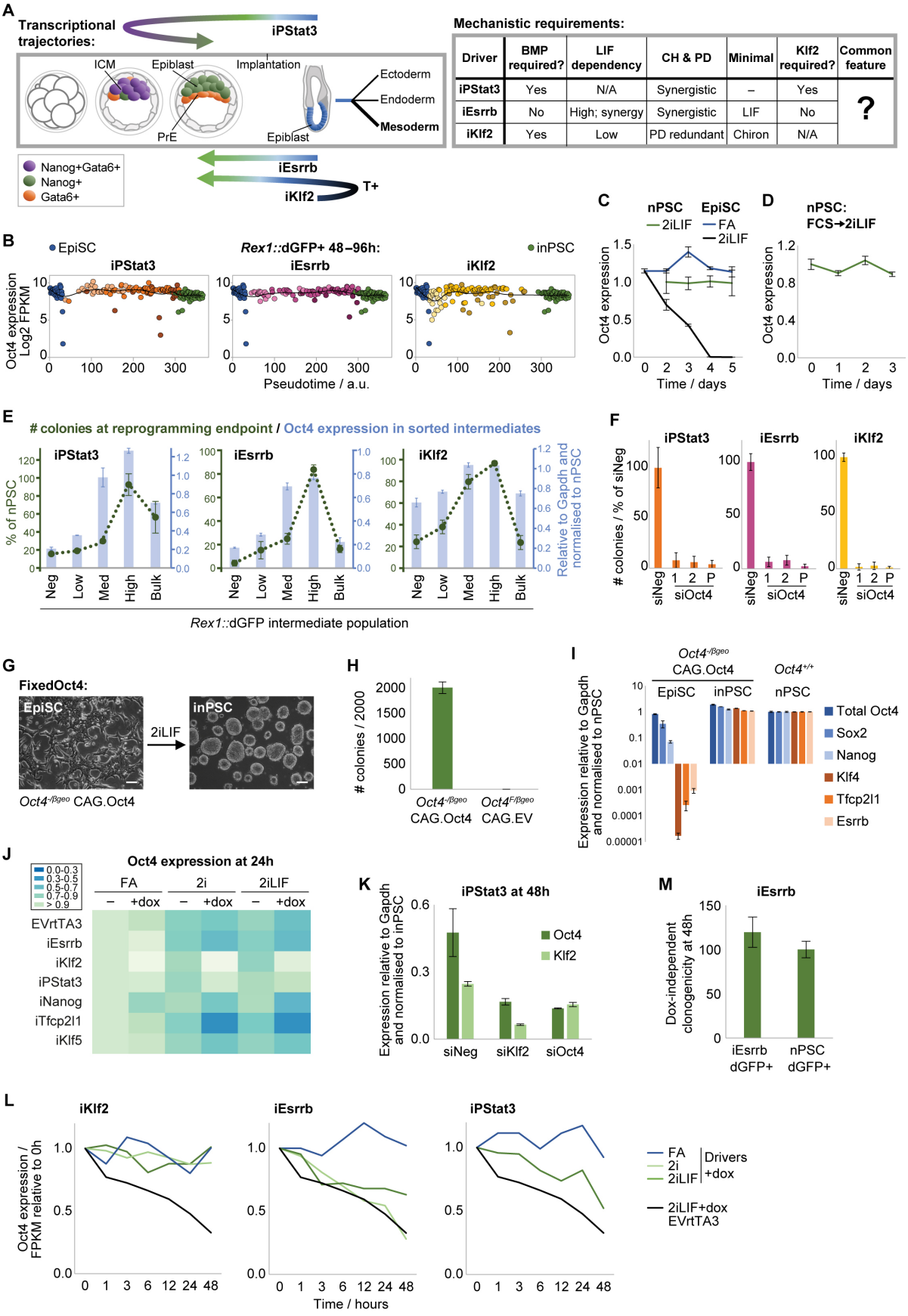


Figure 6: EpiSC reprogramming converges on the fine-tuning of Oct4 expression

(A) Summary of transcriptional trajectories and mechanistic requirements for each driver. **(B)** Scatter plots of Oct4 expression in single cells vs pseudotime (Fig S2C), fitted with LOESS regression lines. **(C–D)** Timecourse RT-qPCR analyses of mean Oct4 expression, displayed relative to Gapdh and normalised to nPSCs ±SD (n=3). **(C)** EVrtTA3 control EpiSCs in FA or 2iLIF+dox, and nPSCs maintained in 2iLIF. **(D)** nPSCs previously cultured in

FCS+LIF then switched to 2iLIF. **(E)** *Rex1::*dGFP negative, low, medium, high and bulk reprogramming intermediates were isolated by flow cytometry, analysed for average Oct4 expression level by RT-qPCR (blue), then replated for clonogenicity assay in 2iLIF (green). Means are presented ±SD (n=3). **(F)** Oct4 KD was performed at reprogramming onset with a single pulse of 10μM siRNA. inPSC colonies scored on day 8 are presented as mean ±SD (n=3). P=Pool. **(G–I)** FixedOct4 EpiSCs formed inPSC colonies at high efficiency in 2iLIF, indicated morphologically (scale 100μm) **(G)**, by mean inPSC colonies scored on day 8 ±SD (n=3) **(H)**, and by RT-qPCR analyses **(I)**. **(J)** Heatmap of Oct4 expression after 24h of 2iLIF+dox/GCSF, measured by RT-qPCR relative to Gapdh then normalised to EpiSCs. **(K)** Oct4 or Klf2 KD was performed at iPStat3 reprogramming onset. After 48h, expressions were analysed by RT-qPCR. **(L)** Timecourse Oct4 expression analyses for EpiSCs induced in the indicated conditions, relative to 0h. **(M)** *Rex1::*dGFP+ iEsrrb reprogramming intermediates (2iLIF+dox) and nPSCs (2iLIF) were isolated by FACS at 48h and plated for clonal assay in 2iLIF without dox. Blasticidin was applied on day 6. Mean naïve colony number scored on day 9 is presented as % of nPSC colonies ±SD (n=3). **See also Figure S6.**

2013. This uncouples Oct4 expression from identity or environmental perturbations, i.e. prevents the loss of Oct4 upon switch of EpiSCs to 2iLIF. An *Oct4* null background was necessary to ensure maintenance of total Oct4 level, and to avoid overexpression of Oct4 that triggers differentiation (Niwa et al., 2000). Correspondingly, ectopic Oct4 expression on top of a wild-type background gives very inefficient EpiSC reprogramming (Guo and Smith, 2010; Yang et al., 2019).

Following medium switch to 2iLIF, FixedOct4 EpiSCs rapidly generated inPSC colonies with extremely high efficiency (Fig 6H–I). The initial transcriptional response of FixedOct4 reprogramming initiation in 2iLIF has aspects in common with each of the other drivers, but is overall most similar to iPStat3 (Fig 1G, S6F). We tested the signal dependencies of FixedOct4 reprogramming, and found that LIF was the minimal requirement for naïve pluripotency induction (Fig S6G–H). In FixedOct4 reprogramming, impetus towards the naïve identity is provided only by exogenous signals: Oct4 is expressed equally in both EpiSCs and inPSCs, so there is no naïve-specific transgene. Therefore, maintenance of Oct4 permits the identity transition, while signals such as LIF specify the direction.

Reconciliation of route differences with common Oct4 maintenance

Despite the distinctions between routes in terms of their transcriptional trajectories and mechanistic requirements (Fig 6A), Oct4 maintenance is a common feature and we further demonstrated that this is required and sufficient for reprogramming (Fig 6F–I). Now we reconcile route-specific attributes with the common denominator of Oct4 maintenance.

First, we assessed the ability of each factor individually to rescue the drop in Oct4 expression when EpiSCs are treated with 2iLIF for 24h (Fig 6J). Klf2 induction yielded the most effective rescue of Oct4 drop, including on the protein level (Fig S6I–K). This Oct4 support could explain the high efficiency of Klf2-driven reprogramming despite its paradoxical dearth of naïve gene induction (Fig 1G). iPStat3 also maintained Oct4 expression (Fig 6J). However, the remaining drivers failed to rescue Oct4 drop in bulk populations.

Since Klf2 is the most effective supporter of Oct4 (Fig 6J) and is an early transcriptional responder to iPStat3 (Fig 5A), we asked whether these observations can be conceptually integrated. Transient Klf2 KD at iPStat3 reprogramming onset resulted in a 65% reduction of Oct4 expression (Fig 6K) and abolished iPStat3-driven reprogramming (Fig 5B). In contrast, Klf2 KD did not abolish reprogramming of FixedOct4 EpiSCs (Fig S6L), even though this is a highly LIF/Stat3-dependent process (Fig S6G–H). This places Oct4 maintenance as a functionally important downstream mechanism of Klf2 in reprogramming, likely to be direct due to its manifestation within 1h (Fig 6L).

iEsrrb was the most efficient of all tested drivers (Fig 1C), yet exhibited an initial drop in Oct4 expression at 24h (Fig 6J, S6J–K) prior to recovery in the productive subpopulation by 48h (Fig 6B, S6B,K). The outstanding feature of iEsrrb reprogramming initiation was the rapid and strong upregulation of naïve genes in a highly 2iLIF-dependent manner (Fig 1G, 5E–F). To test whether this corresponded to rapid wiring of a coherent self-renewing naïve network, we challenged the transgene-independent clonogenicity of iEsrrb *Rex1*::dGFP+ cells at 48h by replating single sorted cells in 2iLIF without dox. Strikingly, their dox-independent clonogenicity was comparable to nPSCs (Fig 6M), indicating that 48h post-induction a functional naïve network has already formed for iEsrrb. Thus

we propose that iEsrrb drives a rapid transition between primed and naïve networks, rescuing Oct4 expression within the 'vulnerable window' between different self-renewing states.

Together, these results indicate that, irrespective of the mechanism used by the different routes, achieving a PSC-level of Oct4 is the common feature of successful reprogramming. This event creates the opportunity for transition into naïve pluripotency, which is effected provided there is a conducive signal environment.

PSC-level of Oct4 is sufficient for somatic cell reprogramming

To address the applicability of our findings to other cellular contexts, we derived somatic cells from FixedOct4 nPSCs by differentiation in chimeras (Fig 7A). Extensive analysis of E9.5 chimera cryosections confirmed bonafide development, with widespread contribution of FixedOct4 cells to all germ lineages, expressing appropriate tissue-specific markers together with Oct4 (Fig 7B, S7A–D). FixedOct4 nPSCs were also capable of performing tetraploid complementation, a stringent assay for developmental contribution (Fig S7E).

Having verified the contribution of FixedOct4 cells to downstream lineages in E9.5 chimeras, we then tested whether they could reprogram and if signal instruction was sufficient. After discarding a generous portion of the tail to stringently avoid germ cell contamination, we dissociated and cultured the anterior portion of each chimera to test reprogramming ability using three different conditions (Fig 7C): directly in 2iLIF, in LIF only, or in LIF combined with a low dose of 5-Azacytidine (aza, an inhibitor of DNA methyltransferase activity, in case assistance was required to remodel a more constrained epigenetic landscape). After 6 days, all conditions were swapped to 2iLIF (Fig 7C). With the exception of positive control allantois, inPSCs were not generated when plated directly in 2iLIF, consistent with our previous demonstration that, if applied from the beginning, 2iLIF does not support somatic cell reprogramming (Silva et al., 2008). However, inPSCs were generated from 16/17 chimeras following culture in LIF+aza, and from 7/9 chimeras after LIF only (Fig 7C–F, S7F). Therefore, LIF is sufficient to induce reprogramming of FixedOct4 cells from E9.5 states, as well as from EpiSCs (Fig S6G–H).

To test more developmentally advanced starting material, we derived FixedOct4 fibroblasts from E12.5 chimeras and investigated whether they could reprogram under signal instruction alone (Fig 7G–J). E12.5 is a standard stage for murine embryonic fibroblast (MEF) derivation as the starting material for somatic cell reprogramming. FixedOct4 MEFs exhibited normal MEF morphology (Fig 7I) and expressed both Oct4 and MEF markers (Fig 7J). Since MEF reprogramming usually takes longer than from EpiSCs and has different signal requirements in the early stages, we tried various conditions in the first week (Fig 7G–H). On day 7, we swapped all conditions to 2iLIF, then on day 14 applied G418 to select for inPSC colonies. As expected, direct application of 2iLIF did not yield inPSCs from MEFs but did allow derivation of naïve pluripotent colonies from genital ridges (positive control). Unlike from FixedOct4 EpiSCs and E9.5 cells, LIF alone was insufficient to reprogram MEFs. Since MEFs may not effectively transduce LIF signal, we added IL6 and soluble IL6 receptor to assist Jak/Stat pathway activation. We also tested addition of FGF2 and Chiron, since there is precedent for a positive role of these signals in fibroblast reprogramming (Giulitti et al., 2019; Li et al., 2011). We successfully obtained inPSCs from IL6+IL6R+LIF+FGF2+Chiron (ILFC), and IL6+IL6R+LIF (IL) (Fig 7H–J). Whilst ILFC was more efficient, IL represents the minimum requirement for MEF reprogramming.

Together, this defines fine-tuned Oct4 expression together with Jak/Stat signalling as sufficient for naïve pluripotency induction from a range of cell types: EpiSCs, E9.5 cells, and E12.5 MEFs.

Discussion

We show that there are multiple routes by which the naïve pluripotent identity can be established from EpiSCs, with the unifying feature of active Oct4 maintenance. Not only do these routes differ in their transcriptional trajectories, but crucially also in their mechanistic attributes of genetic and signal requirements (Fig 6A). Nevertheless, molecular and functional equivalency of the resultant inPSCs demonstrates that these routes ultimately converge to a single identity (Fig 1E–F). Thus, there is considerable flexibility for

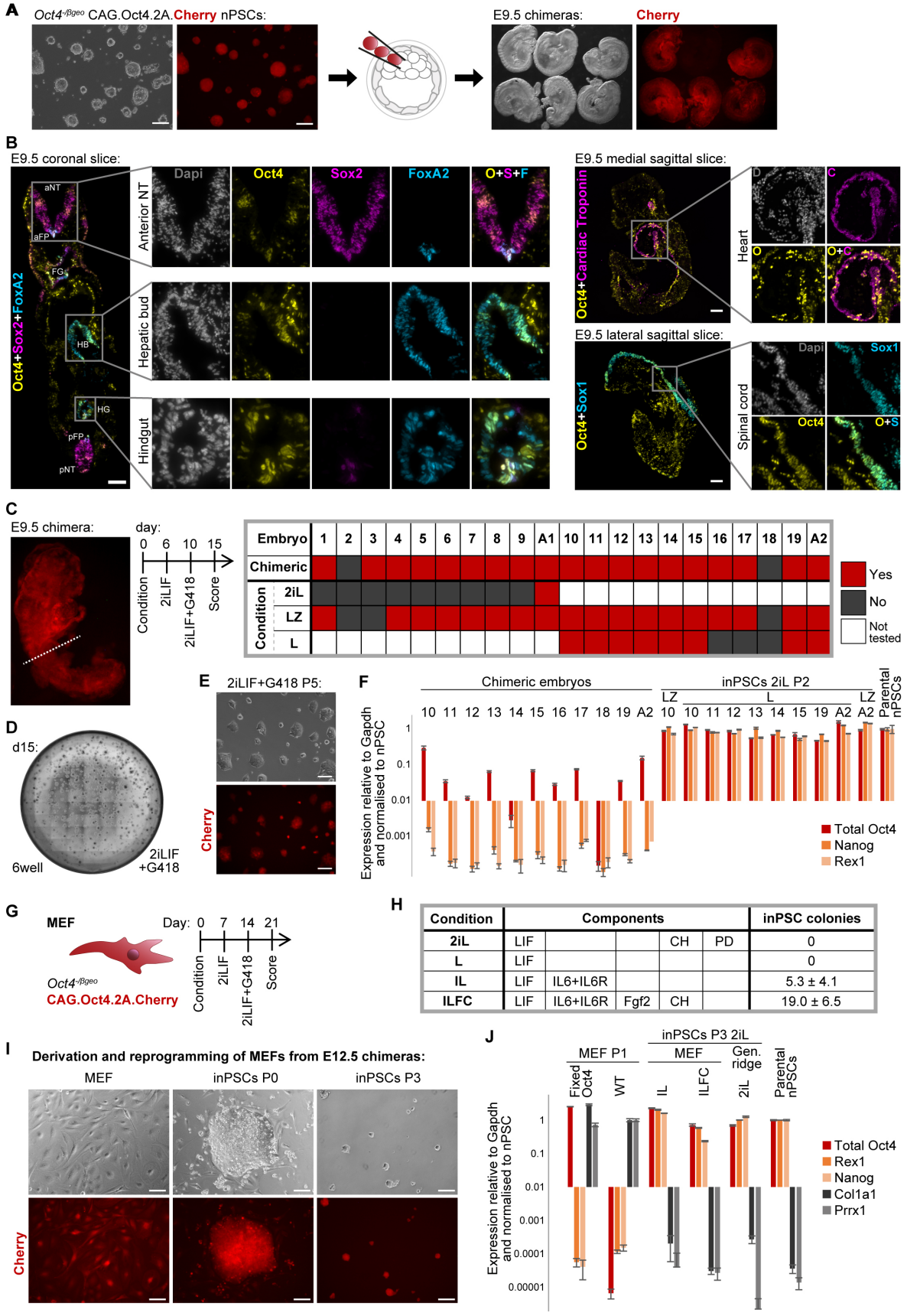


Figure 7: PSC-level of Oct4 is sufficient for somatic cell reprogramming

(A) FixedOct4 nPSCs were injected into E3.5 C57BL/6 blastocysts then transferred to recipients. Resultant embryos were collected at E9.5. Images are shown of 5 chimeras & 1 negative control from the same litter. **(B)** Contribution of FixedOct4 cells to E9.5 chimeras was assessed by immunostaining of 8µm cryosections. Zooms are shown of the indicated regions. Scale: 100µm. NT=neural tube; FP=floor plate; FG=foregut; HB=hepatic bud; HG=hindgut; a=anterior; p=posterior. **(C–J)** Reprogramming of FixedOct4 cells from E9.5 chimeras and E12.5

MEFs. **(C)** Anterior portion of each E9.5 chimera was dissociated manually, subdivided into quarters then cultured in the indicated conditions in duplicates. Generation of inPSCs is summarised in the table. L=LIF; Z=aza; A=allantois. **(D)** inPSCs at P0, following reprogramming of 1/8th of an E9.5 chimera in a 6well. **(E)** inPSCs at P5. Scale: 100µm. **(F)** RT-qPCR analyses of inPSCs at P2, after reprogramming from E9.5 in LIF (L) or LIFaza (LZ), followed by 2iLIF. Mean expressions are shown ± SD (2 technical replicates per embryo). **(G)** Reprogramming protocol for FixedOct4 MEFs, after derivation from E12.5 chimeras. **(H)** Conditions tested during the first week of MEF reprogramming. Number of inPSC colonies scored at day 21 is shown as mean ± SD (n=3), per 5000 MEFs plated. **(I)** FixedOct4 MEFs, an inPSC colony at day 21, and P3 inPSCs. Scale: 100µm. **(J)** RT-qPCR analyses of FixedOct4 and wild-type MEFs, FixedOct4 MEF-derived inPSCs after reprogramming in IL or ILFC followed by 2iLIF, and after derivation directly in 2iLIF for the genital ridge. Mean expressions are shown ± SD (n=3). **See also Figure S7.**

the specification of a single identity from a single origin. This adds further complexity to the paradigm of multicellular biology by which TFs and signals are used in different permutations and contexts in order to generate different cell types: they can also be used in different ways to generate the same cell type.

We relate our reprogramming routes to development by transcriptome comparison, reporter live-imaging, and *in vivo* lineage tracing. iPStat3 intermediates transcriptionally resemble the early embryo ICM and, remarkably, gain its greater developmental potency (Fig 4). In contrast, the iKlf2 route acquires mesodermal signature prior to naïve pluripotency induction (Fig 3). Therefore, initially moving backwards or forwards in developmental time can be compatible with successful reprogramming, provided key mechanistic criteria are met (Fig 6).

Adachi et al., 2018 recently reported that Esrrb acts as a pioneer TF during EpiSC reprogramming, binding to closed chromatin and recruiting the P300 transcriptional coactivator in a LIF/Stat3-dependent manner. This is consistent with our observation that Esrrb-driven reprogramming is highly LIF-dependent (Fig 5C–F). Meanwhile, Stat3 and Smad1 have been reported to form a protein complex together with P300 under conducive signalling conditions (Onishi et al., 2014), which is compatible with our finding that iPStat3-driven reprogramming is blocked by BMP signalling inhibition (Fig 5H). Based on this, we speculate that different reprogramming drivers engage with P300 via different partners, and that this might underpin their different mechanistic requirements (Fig 6A).

iKlf2 is enigmatic as an efficient EpiSC reprogramming driver. Its dearth of naïve gene upregulation within the first 48h is counterintuitive, as is its highly divergent initiation trajectory (Fig 1–3). In the first 48h of reprogramming, the only positive impact on pluripotency genes is robust support of Oct4 expression (Fig 6). Since FixedOct4 is sufficient for highly efficient reprogramming, we reason that a similar phenomenon happens here: iKlf2 intermediates are Oct4+ and thus remain permissive for reprogramming directed by signals. We note that Oct4 is initially maintained during mesendoderm lineage entry (Downs, 2008; Thomson et al., 2011) and thus reason that transient lineage diversion can benefit reprogramming if it helps to achieve the Oct4 maintenance requirement. This signifies a conceptual shift, exposing expression of a transition factor as more important than the transcriptional program directly induced by a driver. Thus, identity change does not simply require activation of the destination program, but instead pivots on the mechanism that permits a transition to occur.

Ultimately, successful reprogramming routes can be thought of as different strategies that converge on the unifying, required and sufficient feature of fine-tuned Oct4 expression (Fig 6). In light of this we propose the following hypothesis: for a given EpiSC reprogramming driver, there is a certain probability of rescuing Oct4 during the ‘vulnerable window’ after Oct4 loses support from the collapsing primed network. We suggest that reprogramming efficiency correlates with this probability, which is determined by the ability of that factor itself to drive Oct4 expression and the speed at which that factor orchestrates a coherent naïve network to support Oct4 in an alternative topology. iKlf2 and iEsrrb occupy opposite extremes within this model, relying solely on the former and latter methods respectively (Fig S7G).

Results from other cell contexts further demonstrate that identity transition to naïve pluripotency pivots on precise Oct4 expression. PSC-level of Oct4 is the minimal requirement for naïve pluripotency induction not only from EpiSCs but also from developmentally more advanced cell types including MEFs (Fig 7). We suggest that correct Oct4 level

creates the opportunity for transition into naïve pluripotency, which is effected provided there is a conducive environment. In agreement with this, Liu et al., 2018 recently reported that CRISPR-based chromatin remodelling of the *Oct4* locus is sufficient to reprogram MEFs, using the acetyltransferase domain of P300 to activate endogenous Oct4. Thus, precise Oct4 expression is the defining feature in distinct contexts of nuclear reprogramming. It will now be interesting to explore how our findings relate to other advances made towards the optimisation and understanding of induced pluripotency.

Whilst Oct4 expression at PSC-level is required and sufficient for reprogramming under signal instruction, it is also compatible with bonafide development when returned to the embryo. In our FixedOct4 system, opposing yet highly efficient identity transitions occur depending solely on the environment: induction of naïve pluripotency in the presence of LIF (Fig 6), or re-entry to development *in vivo* (Fig 7). Oct4 plays a transition-permitting role during early differentiation of several lineages (Niwa et al., 2000; Radzisheuskaya et al., 2013), and can be briefly utilised to promote direct transdifferentiation from fibroblast to neural identity (Thier et al., 2012). In this light, and considering that low-Oct4 traps nPSCs in self-renewal (Karwacki-Neisius et al., 2013; Radzisheuskaya et al., 2013), we now define Oct4 as a 'transition factor', permitting identity change in various directions depending on the context.

Our work supports theories that cell identities are multidimensional attractors, occupying local minima of stable network states (Huang et al., 2005; Kauffman, 1993). Here, we provide a substantial advance on previous works, reaching a single destination identity via three different trajectories. Mechanistic as well as transcriptional differences verify that transitions occur via truly distinct intermediate states. Furthermore, we reveal the logic underpinning multidimensional access to the single attractor: fine-tuned support of a transition factor, in this case Oct4. This provides a conceptual framework for the understanding of cell identity transitions. In the future, it will be of interest to continue identifying the transition factors and supporting logic for the multitude of developmental, regenerative and pathological cell identity transitions.

Author Contributions

HTS and JCRS conceived the study. HTS designed and performed experiments, analysed the data, supervised the study and wrote the manuscript. GGS performed bioinformatic analyses, overseen by PB. TL, LEB, EJS, KM, AM, MRPA, PH and WM performed experiments. MK, AR and RLL created reagents. CYL and SN assisted with scRNAseq, supervised by BG. WR shared data. JN derived cell lines and oversaw embryo experiments. JCRS supervised the study, wrote and approved the manuscript.

Acknowledgements

We thank C Cossetti, S Lees, M Paramor, I Pshenichnaya and A Riddell for specialist technical support. We are grateful to C Handford, K Jones and A Yanagida for technical assistance, and to P Hajkova, M Huch, T Kalkan, C Mulas, A Smith and B Steventon for discussion. HTS is funded by MRC PhD Studentship 1233706, JCRS by Wellcome Fellowship WT101861, and BG by Bloodwise, CRUK, Wellcome and NIH-NIDDK. The authors gratefully acknowledge core support from the Wellcome-MRC Cambridge Stem Cell Institute.

References

Adachi, K., Kopp, W., Wu, G., Heising, S., Greber, B., Stehling, M., Araúzo-Bravo, M.J., Boerno, S.T., Timmermann, B., et al. (2018). Esrrb Unlocks Silenced Enhancers for Reprogramming to Naive Pluripotency. Cell Stem Cell.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat. Cell Biol. 16, 516–528.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature 448.

Buganim, Y., Faddah, D.A., and Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. Nat. Rev. Genet. 14, 427–439.

Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Dev. Biol. 210, 30–43.

Cuny, G.D., Yu, P.B., Laha, J.K., Xing, X., Liu, J.-F., Lai, C.S., Deng, D.Y., Sachidanandan, C., Bloch, K.D., and Peterson, R.T. (2008). Structure–activity relationship study of bone morphogenetic protein (BMP) signaling inhibitors. Bioorg. Med. Chem. Lett. 18, 4388–4392.

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science 343, 193–196.

Downs, K.M. (2008). Systematic localization of oct-3/4 to the gastrulating mouse conceptus suggests manifold roles in mammalian development. Dev. Dyn. 237, 464–475.

Enver, T., Pera, M., Peterson, C., Andrews, P.W., Doble, B., Woodgett, J., Cohen, P., Smith, A., et al. (2009). Stem Cell States, Fates, and the Rules of Attraction. Cell Stem Cell 4, 387–397.

Fehling, H.J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G., and Kouskoff, V. (2003). Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. Development 130, 4217–4227.

Freyer, L., Schröter, C., Saiz, N., Schrode, N., Nowotschin, S., Martinez-Arias, A., and Hadjantonakis, A.-K. (2015). A loss-of-function and H2B-Venus transcriptional reporter allele for Gata6 in mice. BMC Dev. Biol. 15, 38.

Giulitti, S., Pellegrini, M., Zorzan, I., Martini, P., Gagliano, O., Mutarelli, M., Ziller, M.J., Cacchiarelli, D., et al. (2019). Direct generation of human naive induced pluripotent stem cells from somatic cells in microfluidics. Nat. Cell Biol. 21, 275–286.

Graham, S.J.L., Wicher, K.B., Jedrusik, A., Guo, G., Herath, W., Robson, P., and Zernicka-Goetz, M. (2014). BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. Nat. Commun. 5.

Guo, G., and Smith, A. (2010). A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. Development 137, 3185–3192.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development 136.

Hao, J., Ho, J.N., Lewis, J.A., Karim, K.A., Daniels, R.N., Gentry, P.R., Hopkins, C.R., Lindsley, C.W., and Hong, C.C. (2010). In vivo structure - Activity relationship study of dorsomorphin analogues identifies selective VEGF and BMP inhibitors. ACS Chem. Biol. 5, 245–253.

Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D.E. (2005). Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network. Phys. Rev. Lett. 94, 128701.

Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H.J., Nett, I., Marks, H., Walker, R., et al. (2017). Tracking the embryonic stem cell transition from ground state pluripotency. Development 144.

Karwacki-Neisius, V., Göke, J., Osorno, R., Halbritter, F., Ng, J.H., Weiße, A.Y., Wong, F.C.K., Gagliardi, A., et al. (2013). Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. Cell Stem Cell 12.

Kauffman, S.A. (1993). The Origins of Order: self-organisation and selection in evolution.

Li, Y., Zhang, Q., Yin, X., Yang, W., Du, Y., Hou, P., Ge, J., Liu, C., et al. (2011). Generation of iPSCs from mouse fibroblasts with a single gene, Oct4 and small molecules. Cell Res. 21, 196–204.

Liu, P., Chen, M., Liu, Y., Qi, L.S., and Ding, S. (2018). CRISPR-Based Chromatin Remodeling of the Endogenous Oct4 or Sox2 Locus Enables Reprogramming to Pluripotency. Cell Stem Cell.

Martello, G., and Smith, A. (2014). The Nature of Embryonic Stem Cells. Annu. Rev. Cell Dev. Biol. 30, 647–675.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell 11, 491–504.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. 20, 1215–1228.

Morgani, S., Nichols, J., and Hadjantonakis, A.-K. (2017). The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. BMC Dev. Biol. 17, 7.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev 12, 2048–2060.

Niwa, H., Miyazaki, J., Smith, A.G., and Miyazaki, J. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. 24, 372–376.

Onishi, K., Tonge, P.D., Nagy, A., and Zandstra, P.W. (2014). Local BMP-SMAD1 signaling increases LIF receptor-dependent STAT3 responsiveness and primed-to-naive mouse pluripotent stem cell conversion frequency. Stem Cell Reports 3, 156–168.

van Oosten, A.L., Costa, Y., Smith, A., and Silva, J.C. (2012). JAK/STAT3 signalling is sufficient and dominant over antagonistic cues for the establishment of naïve pluripotency. Nat Commun 3, 817.

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A.-K.A.-K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development 135, 3081–3091.

Radzisheuskaya, A., Chia G.L.B., dos Santos, R.L., Theunissen, T.W., Castro, L.F., Nichols, J., and Silva, J.C. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nat Cell Biol 15, 579–590.

Reyes de Mochel, N.S., Luong, M., Chiang, M., Javier, A.L., Luu, E., Toshihiko, F., MacGregor, G.R., Cinquin, O., and Cho, K.W.Y. (2015). BMP signaling is required for cell cleavage in preimplantation-mouse embryos. Dev. Biol. 397, 45–55.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.-K.K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., Wrana, J.L., et al. (2010). Functional genomics reveals a BMP-Driven mesenchymal-to-Epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell 7, 64–77.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol. 6, e253.

Smith, Z.D., Sindhu, C., and Meissner, A. (2016). Molecular features of cellular reprogramming and development. Nat.Rev.Mol.Cell Biol. 17, 139–154.

Stuart, H.T., Van Oosten, A.L., Radzisheuskaya, A., Martello, G., Miller, A., Dietmann, S., Nichols, J., and Silva, J.C.R.J.C.R. (2014). NANOG amplifies STAT3 activation and they synergistically induce the naive pluripotent program. Curr. Biol. 24.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448, 196–199.

Thier, M., Wörsdörfer, P., Lakes, Y.B., Gorris, R., Herms, S., Opitz, T., Seiferling, D., Quandel, T., Hoffmann, P., Nöthen, M.M., et al. (2012). Direct conversion of fibroblasts into stably expandable neural stem cells. Cell Stem Cell 10, 473–479.

Thomson, M., Liu, S.J., Zou, L.-N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell 145, 875–889.

Yang, J., van Oosten, A.L., Theunissen, T.W., Guo, G., Silva, J.C.R.R., and Smith, A. (2010). Stat3 activation is limiting for reprogramming to ground state pluripotency. Cell Stem Cell 7.

Yang, J., Rajan, S.S., Friedrich, M.J., Lan, G., Zou, X., Ponstingl, H., Garyfallos, D.A., Liu, P., Bradley, A., and Metzakopian, E. (2019). Genome-Scale CRISPRa Screen Identifies Novel Factors for Cellular Reprogramming. Stem Cell Reports 12, 757–771.

Yeo, J.-C., Jiang, J., Tan, Z.-Y., Yim, G.-R., Ng, J.-H., Göke, J., Kraus, P., Liang, H., et al. (2014). Klf2 Is an Essential Factor that Sustains Ground State Pluripotency. Cell Stem Cell 14, 864–872.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development 122.

Ying, Q.-L., Nichols, J., Chambers, I., Smith, A., Adelman, C.., Chattopadhyay, S., Bieker, J.., Attisano, L., Wrana, J., et al. (2003). BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. Cell 115, 281–292.

Ying, Q.-L.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature 453, 519–523.

Supplemental Information

Distinct routes converge on a unifying transition logic to establish naïve pluripotency

Hannah T Stuart, Giuliano G Stirparo, Tim Lohoff, Lawrence E Bates, Masaki Kinoshita, Chee Y Lim, Elsa J Sousa, Katsiaryna Maskalenka, Aliaksandra Radzisheuskaya, Andrew Malcolm, Mariana R P Alves, Rebecca L Lloyd, Sonia Nestorowa, Peter Humphreys, William Mansfield, Wolf Reik, Paul Bertone, Jennifer Nichols, Berthold Göttgens, José C R Silva

Supplemental Figure Legends

Figure S1: Reprogramming initiation is driver-dependent
(A) Gene expression of established inPSCs at passage 5, together with control EpiSCs and nPSCs. Means are shown from scRNAseq. Greys: primed markers. Blues: core pluripotency markers. Oranges: naïve markers. Together with Fig 1E, this shows that inPSCs derived with each driver do not differ in their molecular signatures.
(B) Percentage of *Rex1::*dGFP+ cells is shown for each driver after reprogramming induction in 2iLIF+dox (iEsrrb, iKlf2) or 2iLIF+GCSF (iPStat3), from days 2–5. Note that in standard reprogramming assays we would normally withdraw dox/GCSF at day 4 and add blasticidin to select for *Rex1* reporter activity. However, in this instance we continued in 2iLIF+dox/GCSF until day 5 and never added blasticidin, so that the % of dGFP+ cells was not confounded. Parental *Rex1::*dGFP EpiSCs in FA (0.00% dGFP+) and *Rex1::*dGFP nPSCs in 2iLIF (99.50% dGFP+) provided negative and positive controls respectively. Example: FACS-plot showing emergence of *Rex1::*dGFP expression during reprogramming of iEsrrb in 2iLIF+dox at day 4.
(C) Reprogramming of iEsrrb, iPStat3 and iKlf2 EpiSCs was induced at day 0 in 2iLIF+dox (iEsrrb, iKlf2) or 2iLIF+GCSF (iPStat3). Transgene induction by dox/GCSF was withdrawn on either day 1, 2, 3 or 4, with concomitant addition of blasticidin (bsd) to select for *Rex1* reporter activity. Naïve colonies were scored on day 8, and are presented as mean ±SD (n=3), relative to day 4. iEsrrb is the fastest to yield transgene-independent inPSCs, while iKlf2 is the slowest, consistent with their differing rates of *Rex1::*dGFP induction (Fig S1B) and their differing kinetics of naïve gene expression induction (Fig 1G).
(D) Timecourse of driver protein induction from 0–48h. Western blots are shown against PStat3, Klf2 and Esrrb, with αTubulin providing loading control. Induction is robust from 1h onwards for all drivers. Therefore, differences in naïve network induction kinetics are due to the downstream responses, rather than due to delays in transgene induction itself.
(E) Immunofluorescent staining was performed and quantified 24h after transgene induction, on a total of 5956 cells. EpiSC and nPSC samples provide negative and positive controls respectively. To determine the % of driver-positive cells, a stringent threshold was calculated: the mean of EpiSC values plus three standard deviations (3SD), indicated on the plots. Quantification of driver proteins in single cells shows efficient inductions (93–98%) to expression levels comparable to those of the endogenous proteins in nPSCs, which is the biologically relevant reference. DAPI+ area is plotted on the *y*-axis simply to assist in data visualisation.
(F) RNAseq was performed on bulk samples after driver induction in 2i±LIF and in FA. Principle component analyses (PCA) based on expressed genes (FPKM>0) are shown for iEsrrb (above) and iKlf2 (below).
(G) *k*-means clustering of bulk RNAseq samples, based on expressed genes (FPKM>0). Optimal *k* cluster number was computed using the elbow method.

Figure S2: Single-cell RNAseq defines distinct productive trajectories
(A) Unsupervised hierarchical clustering, computed based on all genes with the Ward.D2 agglomeration method and Euclidean distances. iKlf2 12/24h cells cluster in a separate branch '*a*' from all other samples, indicating that they are more different to the rest than even EpiSC (start) vs nPSC/inPSC (end) samples are to each other. This is consistent with an initial diversion in the iKlf2 trajectory. The remainder of the samples clustered into indicated branches *b* and *c*. Branch *b* subclustered into: *b(i)*, comprised of EpiSCs and mostly 12h iPStat3 and iEsrrb; *b(ii),* mostly 24h iPStat3 and iEsrrb; *b(iii),* a mixture of early and mostly 72h dGFP- samples. Branch *c* subclustered into: *c(i)*, comprised of nPSC/inPSCs; and *c(ii)*, mostly 72h dGFPlow and dGFP+ samples for all drivers. Overall, this indicates that early timepoints and less productive populations are more similar to EpiSCs (start identity) whereas later and higher dGFP samples are more similar to nPSC/inPSCs (end identity). Importantly, within groups *b* and *c*, cells cluster according to driver rather than according to timepoint or dGFP status, demonstrating that routes are transcriptionally distinct throughout.
(B) Differential expression (DE) analysis was performed on each sample set relative to start EpiSCs. The resulting lists were compared and Venn diagrams plotted to find the numbers of

DE genes that are unique to or shared between drivers at each timepoint. Examples of the Venn diagrams are shown here, and the DE numbers are summarised in main Fig 2E.
(C) Computation of fraction of similarity between each single cell vs EpiSCs in FA and inPSCs in 2iLIF. Signature EpiSC and inPSC datasets were generated by averaging of bulk RNAseq samples. Single cells were ordered from lowest to highest inPSC/EpiSC identity fractions to generate pseudotime coordinates from reprogramming start to end. Pseudotime coordinates largely agreed with real-time.
(D) Scatter plots of expression in single cells vs pseudotime, fitted with LOESS regression lines.

Figure S3: iKlf2 reprogramming proceeds via a mesoderm-like state
(A) Scatter plots of mesodermal markers (Fgf8, Mesp1, Lefty1, T) against Nanog expression in iEsrrb and iPStat3 reprogramming intermediates, EpiSC (start) and nPSC/inPSC (end) identity controls. The *y*-axes are to the same scales as Fig 3A to facilitate comparison.
(B) Upper: Knock-in strategy for the *Rex1*::mKO2 fusion cassette, which was constructed by replacing the dGFP cassette of the *Rex1*::dGFP targeting vector (Kalkan et al., 2017). E4/5 = exon 4/5; H = helical linker (Arai et al., 2001); mKO2 = monomeric Kusabira Orange 2; bsd confers resistance to blasticidin if *Rex1* is expressed. Lower: resulting bsd-resistant clones were genotyped by PCR using the indicated F1/R1 or F2/R2 primers. TG is the $T^{+/GFP}$ parental nPSC line (Fehling et al., 2003). TGHRO are targeted with the helical linker (H) construct whereas TGGRO clones contained a glycine-serine linker. Correct targeting was obtained in nPSC clones TGHRO6&8, and TGHRO6 was used for subsequent experiments (simply denoted as TGRO in main figure panels).
(C) Immunofluorescent staining against GFP and T proteins, following induction of iKlf2 in TGHRO6 (TGRO) EpiSCs. Scale bars: 100μm.

Figure S4: iPStat3 reprogramming proceeds via an early ICM-like state
(A) Scatter plots of Gata6 vs Nanog expression in iEsrrb and iKlf2 reprogramming intermediates, EpiSC (start) and nPSC/inPSC (end) identity controls. The *y*-axes are to the same scales as Fig 4C to facilitate comparison.
(B) Scatter plots of Gata6 vs earlier or later ICM markers in iPStat3 reprogramming intermediates, E3.5 and E4.5 embryos. ICM: inner cell mass. Epi: epiblast. PrE: primitive endoderm. There is a temporal sequence of naïve gene activation in the embryo (Boroviak et al., 2015). Earlier markers (Klf2, Tfcp2l1) are expressed in Nanog+Gata6+ early ICM. In contrast, later markers (Gbx2, Nr0b1) are not activated until after Nanog+Gata6- naïve epiblast has segregated from Nanog-Gata6+ primitive endoderm. iPStat3 reprogramming intermediates emulate this *in vivo* progression: Klf2 and Tfcp2l1 are turned on earlier and are co-expressed with Gata6, whereas Gbx2 and Nr0b1 are activated later in Gata6- cells.
(C) iPStat3 was induced to reprogram *Gata6::*H2BVenus EpiSCs. Immunofluorescent staining against Gata6 and Nanog was performed at day 2.5. H2BVenus signal persisted without need for counterstaining. Scale bars: 20μm.
(D) Immunofluorescent staining against Gata4 and Sox2, 2.5 days after iPStat3 reprogramming induction, i.e. at the timepoint of sorting for injection. Scale bars: 20μm.

Figure S5: Routes have distinct genetic and signal requirements
(A) Western blots against Esrrb and Klf2, after siRNA treatment of nPSCs in 2iLIF. αTubulin provides loading control. siRNAs were applied from 0–16h, and samples harvested at 48h. Individual siRNAs were used at 10μM, whereas pools were comprised of 4x 2.5μM.
(B) siRNA treatment was performed in a single 16h pulse at reprogramming onset, with individual 10μM siRNAs as well as pools (P) of 4x 2.5μM. Reprogramming was induced with 2iLIF+dox (iEsrrb, iKlf2) or 2iLIF+GCSF (iPStat3). Selection was performed from day 4 with 2iLIF+blasticidin and inPSC colonies were scored at day 8, presented as mean ± SD (n=6) relative to siNeg. Pools of 4 siRNAs, each at a quarter concentration and targeting a different region of the mRNA, are expected to minimise off-target effects. With the exception of inefficient Klf2 siRNA #4 (Fig S5A), individual siRNAs give the same outcomes as pools: iPStat3-

driven reprogramming is dependent on early Klf2 expression, whereas iEsrrb is not (Fig 5B); Esrrb KD at reprogramming onset does not abolish reprogramming by any driver. This is consistent with lack of Esrrb expression by any other driver at 24h (Fig 5A, S5D).
(C) Klf2 KD was performed at iPStat3-driven reprogramming onset with a single 16h pulse of siRNA pool, in the indicated conditions +GCSF from day 0–4, then inPSC colonies were selected from day 4–8 in 2iLIF+blasticidin. inPSC colonies were scored on day 8, presented as mean ±SD (n=3) relative to siNeg. PD=PD03; CH=Chiron; 2i=PD+CH. iPStat3 sensitivity to Klf2 KD is alleviated in the absence of PD03, reminiscent of the ability to rescue embryo-derived *Nanog*$^{-/-}$ EpiSC reprogramming in Chiron+LIF but not 2iLIF (Stuart et al., 2014). Notably, both Nanog and Klf2 are considered targets of PD03 (Silva et al., 2009; Yeo et al., 2014): it appears they are required to transduce its positive input to the naïve network and, in their absence, PD03 is actively detrimental to reprogramming. Hence, interplay between TFs and signals can fundamentally modulate each other's role in identity specification.
(D) Western blots against PStat3 and Esrrb, after 24h in FA or 2iLIF+dox/GCSF. αTubulin provides loading control.
(E) RT-qPCR analysis of EVrtTA3 EpiSCs in the indicated condition +dox. Mean gene expression is displayed relative to Gapdh and normalised to inPSCs, ±SD (n=3). The axes are to the same scales as Fig 5F to facilitate comparison.
(F) Immunofluorescent staining 48h after reprogramming induction. nPSCs cultured in FCS+LIF provide a positive control for PSmad1/5. Maximum intensity projections of Z-stack slices are presented. Scale bars: 20μm.
(G) Validation of 3μM DMH2 efficacy. Wild-type nPSCs cultured in FCS+LIF were treated for 24h as indicated. BMP signalling is known to be active and important for pluripotency maintenance in FCS+LIF (Ying et al., 2003), unlike in 2iLIF. Maximum intensity projections of Z-stack slices are presented. Scale bars: 20μm.
(H) 3μM DMH2 or 1:1000 DMSO was applied to nPSCs or previously established iKlf2 inPSCs, at clonal density in 2iLIF. Naïve colonies were scored after 4 days, as mean ±SD (n=3).
(I–K) Mouse blastocysts were cultured with DMSO or BMP signalling inhibitor from cavitation onset until the late blastocyst stage, by which point the epiblast (Epi) and primitive endoderm (PrE) lineages are fully segregated, thus allowing us to study the impact on each lineage.
(I) Quantitative nPSC-derivation was performed from late blastocysts, following treatment with either DMSO or 3μM DMH2. Immunosurgery was performed to remove the trophectoderm (TE), then the inner cell mass (ICM, comprising Epi+PrE) was dissociated to single cells. 10 single cells were manually transferred to each 96well and cultured in feeder-free 2iLIF conditions, without any further DMSO/DMH2 treatment so that we could assess whether the Epi of the embryo was already affected. By plating 10 single cells per well, we could measure absolute nPSC-derivation efficiency independently of embryo size. The number of nPSC colonies was scored on day 6, and is presented per 10 single cells (main Fig 5K) and as mean per embryo ±SD (here). DMSO n=7; DMH2 n=8. nPSC-identity was subsequently confirmed by RT-qPCR (data not shown).
(J) Mean number of total cells per embryo ± SD. DMSO n=23, DMH2 n=18, LDN n=7 embryos. Note that the lower cell number per DMH2-treated embryo was not due to developmental retardation: Epi and PrE endoderm lineages segregated, the proportion of cells per lineage was unchanged (Fig 5L), and expression of Gata4 indicates late PrE (Artus et al., 2011).
(K) Late blastocysts were fixed and stained for Cdx2, Gata4 and Oct4, following treatment with 1:1000 DMSO, 3μM DMH2 or 0.3μM LDN. Mean signal intensity was quantified for each nucleus, and is presented as box-and-whisker plots for Cdx2 in TE cells (DMSO n=23, DMH2 n=18, LDN n=7 embryos). Oct4 and Gata4 results are presented in main Fig 5N.
(L) A subset of embryos was also stained for Nanog. Mean signal intensity was quantified for each nucleus, and is presented as box-and-whisker plots for Nanog in Epi cells (DMSO n=9, DMH2 n=11, LDN n=4).

Figure S6: EpiSC reprogramming converges on the fine-tuning of Oct4 expression
(A) Immunofluorescent staining of Oct4 expression in *Rex1::*dGFP EpiSCs and inPSCs. Maximum intensity projections of Z-stack slices are presented. Scale bars: 20μm.

(B) Timecourse RT-qPCR analyses of *Rex1::*dGFP+ and dGFP- EpiSC reprogramming intermediates in 2iLIF+dox/GCSF, and of *Rex1::*dGFP+ nPSCs in 2iLIF. Mean Oct4 expression is displayed relative to Gapdh and normalised to nPSC day 2, ±SD (n=3).
(C) FACS-plot showing *Rex1::*dGFP levels during reprogramming. The example here is iEsrrb 2iLIF+dox day 3. *Rex1::*dGFP high, medium (med), low, and negative (neg) gates are indicated. *Rex1::*dGFP negative and high gates were set according to *Rex1::*dGFP EpiSCs and nPSCs respectively. Low and medium gates subdivide the intervening levels.
(D) Western blot against Oct4, after siRNA treatment of nPSCs in 2iLIF. αTubulin provides loading control. siRNAs were applied from 0–16h, and samples harvested at 48h. Individual siRNAs were used at 10μM, whereas pool was comprised of 4x 2.5μM. Whilst other genetic requirements at reprogramming onset are driver-specific (Fig S5B), Oct4 KD at reprogramming onset abolished reprogramming for iKlf2, iEsrrb and iPStat3 (Fig 6F).
(E) Genotyping of *Oct4*$^{-/\beta geo}$ CAG.Oct4 EpiSCs (FixedOct4), with *Oct4*$^{+/+}$ and *Oct4*$^{F/\beta geo}$ controls. *F*=flox allele. The *βgeo* allele confers resistance to G418 if *Oct4* promoter is active.
(F) Gene expression analyses by RT-qPCR of FixedOct4 EpiSCs following reprogramming induction by 2iLIF. Mean gene expression is displayed relative to Gapdh and normalised to inPSCs, ±SD (n=3). The initial transcriptional response of FixedOct4 to 2iLIF has features in common with each of the other drivers: rapid upregulation of Klf2 (as also observed for iPStat3 & iEsrrb); moderate upregulation of Tfcp2l1 (iPStat3 & iEsrrb); poor upregulation of Klf4 (iPStat3 & iKlf2); poor upregulation of Klf5 (iEsrrb & iKlf2); poor upregulation of iEsrrb (iPStat3 & iKlf2). Thus, besides the Klf5 discrepancy, FixedOct4 initiation is most similar to iPStat3, consistent with its reprogramming impetus coming from the environment including LIF.
(G–H) *Oct4*$^{-/\beta geo}$ CAG.Oct4 EpiSCs were plated at 2000/24well in FA. The following day, medium was changed to N2B27 ± Chiron ± PD03 (PD) ± LIF in all permutations (n=3). After 4 days, G418 was applied to select for endogenous *Oct4* promoter activity. Quantification of reprogramming efficiency was inappropriate at this point, since some conditions are permissive compared to 2iLIF. Instead, on day 8, each 24well was passaged in its entirety to 2 x 6wells, one maintaining the condition +G418 and one swapped to 2iLIF+G418 to challenge inPSC clonogenicity in this naïve-selective condition. P1 inPSC colonies were scored on day 12 in 2iLIF+G418, presented as mean ± SD (n=3) **(G).** Condition +G418 phase images are shown for those conditions which successfully generated inPSCs **(H)**. Scale bars: 100μm. We found that FixedOct4 +LIF was the minimal requirement for naïve pluripotency specification, with derivative inPSCs expandable in LIF+G418 or 2iLIF+G418 for at least 8 passages. Curiously, Chiron+LIF did not instruct naïve pluripotency acquisition for FixedOct4 EpiSCs despite initial emergence of naïve-like morphology.
(I) Western blot against Oct4 in EVrtTA3 and iKlf2 EpiSCs, after 24h in FA or 2iLIF ± dox. αTubulin provides loading control.
(J) Immunofluorescent staining of Oct4 expression in *Rex1::*dGFP iKlf2, iPStat3, iEsrrb and EVrtTA3 EpiSCs, after 24h in 2iLIF+dox/GCSF. Scale bars: 100μm.
(K) Quantification of Oct4 immunofluorescent staining during EpiSC reprogramming, on a total of 14,736 single cells. Samples included all cells and were not sorted according to *Rex1::*dGFP reporter, i.e. productive and unproductive cells are present to capture all events unbiasedly. Tfcp2l1 co-staining indicates progression towards the naïve pluripotent identity. Oct4 protein is lost in negative control EVrtTA3 EpiSCs following medium switch, but this can be rescued by the reprogramming drivers. Oct4 is maintained on the protein level in cells progressing towards naïve pluripotency. Conversely, there is no evidence of naïve acquisition if Oct4 protein is lost.
(L) To address the role of endogenous Klf2, Esrrb and PStat3 during FixedOct4 EpiSC reprogramming, Klf2 and Esrrb KD were performed at reprogramming onset with a single 16h pulse of individual (10μM) or pooled (4x 2.5μM) siRNAs. Reprogramming was induced with 2iLIF, selection was performed from day 4 with 2iLIF+G418, then inPSC colonies were scored at day 8, presented as mean ± SD (n=6) relative to siNeg. To address the role of PStat3, we instead conducted experiments in the absence of LIF (L) and/or in the presence of Jak inhibitor (Ji), to ensure that Stat3 was not activated. Overall, LIF/Stat3 signal inhibition had the greatest impact on FixedOct4 reprogramming, whilst Esrrb KD had the least.

Figure S7: PSC-level of Oct4 is sufficient for somatic cell reprogramming
(A–D) Differentiation of FixedOct4 cells in E9.5 chimeras was analysed by immunostaining of 8μm cryosections, taken in various sectioning planes. This extensively demonstrates bonafide contribution of FixedOct4 cells to downstream lineages in the embryo, as evidenced by continued Oct4 expression together with appropriate lineage markers.
(A) Cherry and brightfield (BF) images were acquired on rehydrated sagittal slice, prior to immunofluorescent staining against Oct4 and Nanog. Contribution of FixedOct4 cells was high and widespread, visualised by Cherry signal from the CAG.Oct4.2A.Cherry transgene. Importantly, Cherry signal agreed with Oct4 counter-stain, whereas Nanog expression was not detected. This confirms that the cells had exited pluripotency yet maintained transgenic Oct4 expression. Subsequently, H&E staining was performed on the same slice. Contribution of FixedOct4 cells to various tissues is evident from H&E histological analyses. Scale bars: 200μm.
(B) Immunofluorescent staining against Oct4 and Sox1 on sagittal slice (midline section). Zoom of the indicated region shows the developing brain. Scale bars: 100μm.
(C) Immunofluorescent staining against Oct4 and T on coronal slice. Zoom of the indicated region shows the notochord. Scale bars: 100μm. GE=gut endoderm; NT=neural tube; a=anterior; p=posterior.
(D) Immunofluorescent staining against Oct4 and Sox17 on a sagittal slice. Zoom of the indicated region shows intersomitic blood vessels. Scale bars: 100μm.
(E) Tetraploid C57BL/6 embryos were generated by cell fusion at the 2-cell stage, then cultured to the 8-cell stage. *Oct4*$^{-/\beta geo}$ CAG.Oct4.2A.Cherry nPSCs were injected into tetraploid 8-cell embryos, then transferred to recipients. Resultant embryos were collected at E8.5. Phase and Cherry images are shown, of 3 tetraploid complementations and 3 stage-matched wild-type embryos from a different litter. Scale bars: 100μm. This shows that FixedOct4 cells are capable of performing tetraploid complementation, a stringent assay for developmental contribution.
(F) Differentiated FixedOct4 cells were derived from E9.5 chimeras, then reprogrammed in LIFaza (LZ) or LIF only (L) from day 0–6, followed by 2iLIF until day 10 (Fig 7C). RT-qPCR expression analyses are shown after two days of treatment with L or LZ on cells from chimera 1, for inPSCs at passage 2 in 2iLIF+G418 after derivation in LZ then 2iLIF from chimeras 1–7, and after derivation directly in 2iLIF for a chimeric allantois (A1). Mean expressions ± SD (2 technical replicates per embryo) are presented relative to Gapdh then normalised to parental nPSCs.
(G) Schematic summarising the unifying, required and sufficient feature of correct Oct4 level, which permits cells to transit into naïve pluripotency. Left: diverse logics by which different EpiSC reprogramming drivers achieve correct Oct4 expression. When control EpiSCs are exposed to naïve signals, Oct4 expression drops as the primed network is disrupted. Reprogramming drivers must actively overcome this in order to undergo the identity transition. Although iKlf2, iPStat3 and iEsrrb drive reprogramming by transcriptionally and mechanistically distinct routes (Fig 6A), all ultimately achieve the convergent feature of correct Oct4 level and thus can reach the same naïve pluripotent destination. Right: the functional importance of precise Oct4 expression was confirmed by transient KD, which abolishes reprogramming by all drivers. Conversely, fixing Oct4 to PSC level is sufficient for reprogramming under only signal instruction, from EpiSCs and from developmentally more advanced cell types including MEFs. Therefore, appropriate Oct4 expression is the pivotal feature for transition into the naïve pluripotent identity, regardless of the route of approach.

Figure S1

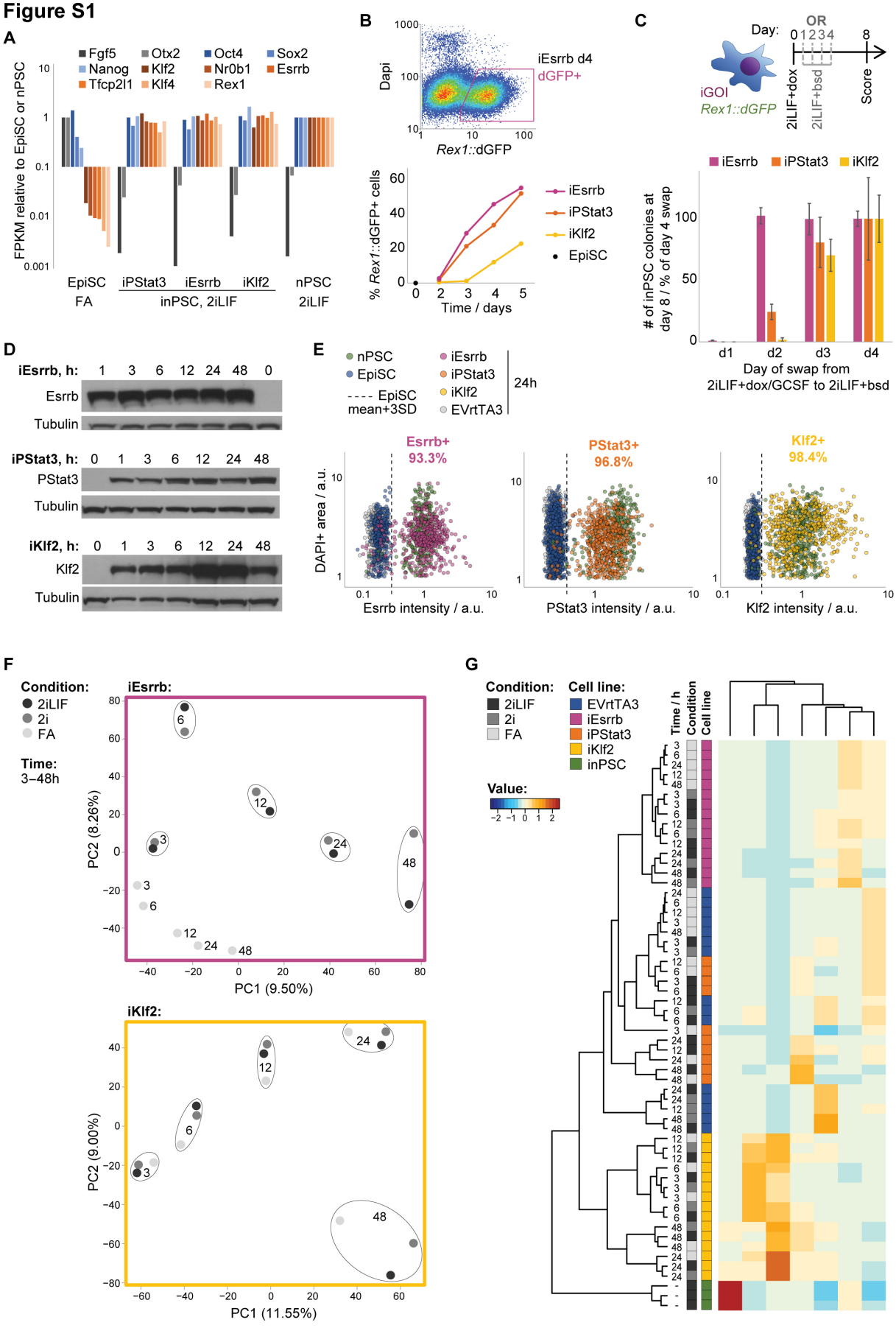

A
Fgf5
Otx2
Oct4
Sox2
Nanog
Klf2
Nr0b1
Esrrb
Tfcp2l1
Klf4
Rex1
FPKM relative to EpiSC or nPSC
EpiSC
iPStat3
iEsrrb
iKlf2
nPSC
FA
inPSC, 2iLIF
2iLIF
B
Dapi
iEsrrb d4
dGFP+
Rex1::dGFP
% Rex1::dGFP+ cells
Time / days
iEsrrb
iPStat3
iKlf2
EpiSC
C
Day:
OR
iGOI
Rex1::dGFP
2iLIF+dox
2iLIF+bsd
Score
# of inPSC colonies at day 8 / % of day 4 swap
Day of swap from 2iLIF+dox/GCSF to 2iLIF+bsd
D
iEsrrb, h:
Esrrb
Tubulin
iPStat3, h:
PStat3
iKlf2, h:
Klf2
E
nPSC
EpiSC
EpiSC mean+3SD
EVrtTA3
24h
Esrrb+ 93.3%
PStat3+ 96.8%
Klf2+ 98.4%
DAPI+ area / a.u.
Esrrb intensity / a.u.
PStat3 intensity / a.u.
Klf2 intensity / a.u.
F
Condition:
2iLIF
2i
FA
Time:
3–48h
iEsrrb:
PC2 (8.26%)
PC1 (9.50%)
iKlf2:
PC2 (9.00%)
PC1 (11.55%)
G
Condition:
Cell line:
EVrtTA3
iEsrrb
iPStat3
iKlf2
inPSC
Value:
Time / h
Condition
Cell line

Figure S2

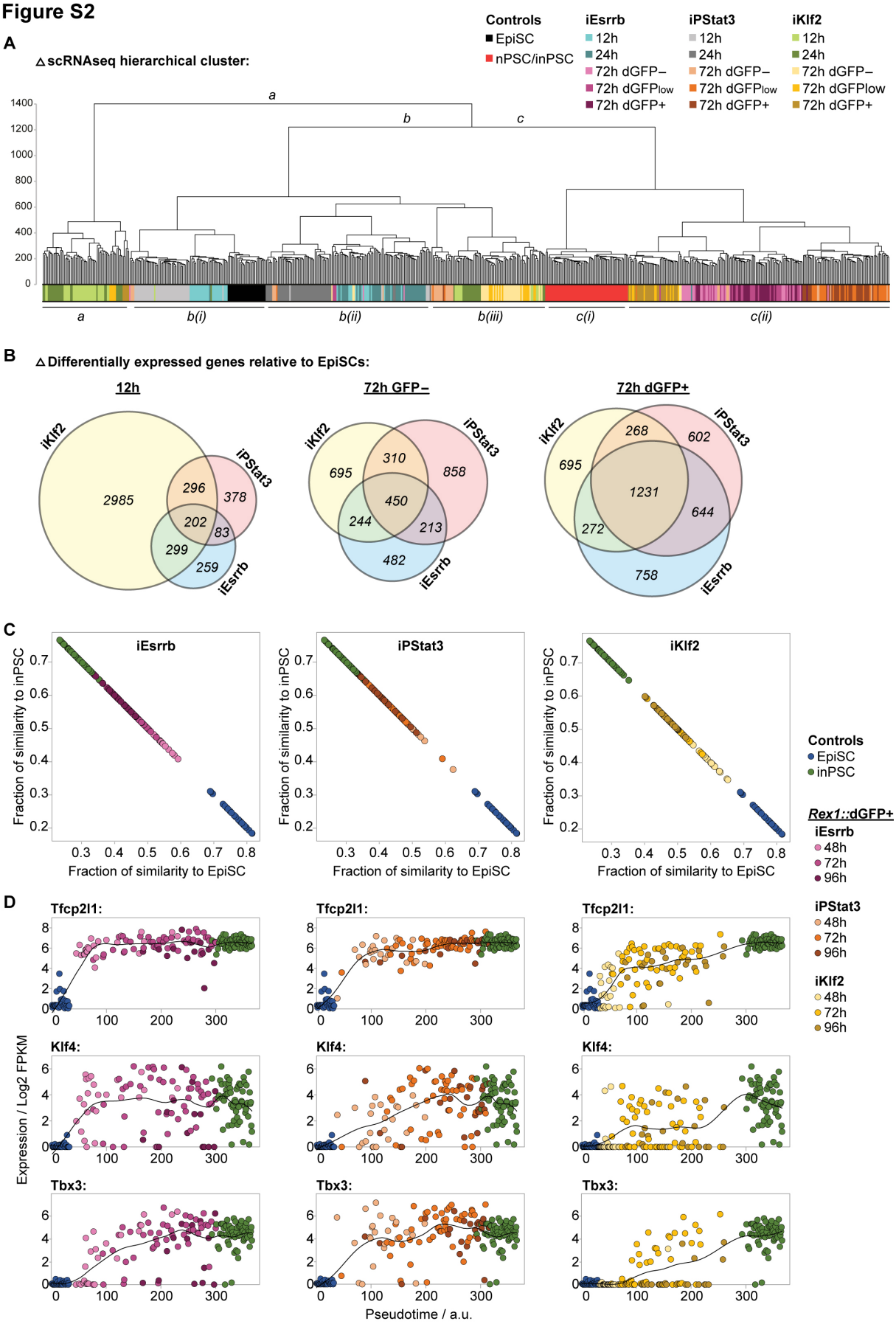

A
△scRNAseq hierarchical cluster:
Controls
EpiSC
nPSC/inPSC
iEsrrb
12h
24h
72h dGFP–
72h dGFPlow
72h dGFP+
iPStat3
12h
24h
72h dGFP–
72h dGFPlow
72h dGFP+
iKlf2
12h
24h
72h dGFP–
72h dGFPlow
72h dGFP+
a
b
c
b(i)
b(ii)
b(iii)
c(i)
c(ii)
B
△Differentially expressed genes relative to EpiSCs:
12h
72h GFP–
72h dGFP+
iKlf2
iPStat3
iEsrrb
2985
296
378
202
83
299
259
695
310
858
450
244
213
482
268
602
1231
644
272
758
C
iEsrrb
iPStat3
iKlf2
Fraction of similarity to inPSC
Fraction of similarity to EpiSC
Controls
EpiSC
inPSC
Rex1::dGFP+
iEsrrb
48h
72h
96h
iPStat3
48h
72h
96h
iKlf2
48h
72h
96h
D
Tfcp2l1:
Klf4:
Tbx3:
Expression / Log2 FPKM
Pseudotime / a.u.

Figure S3

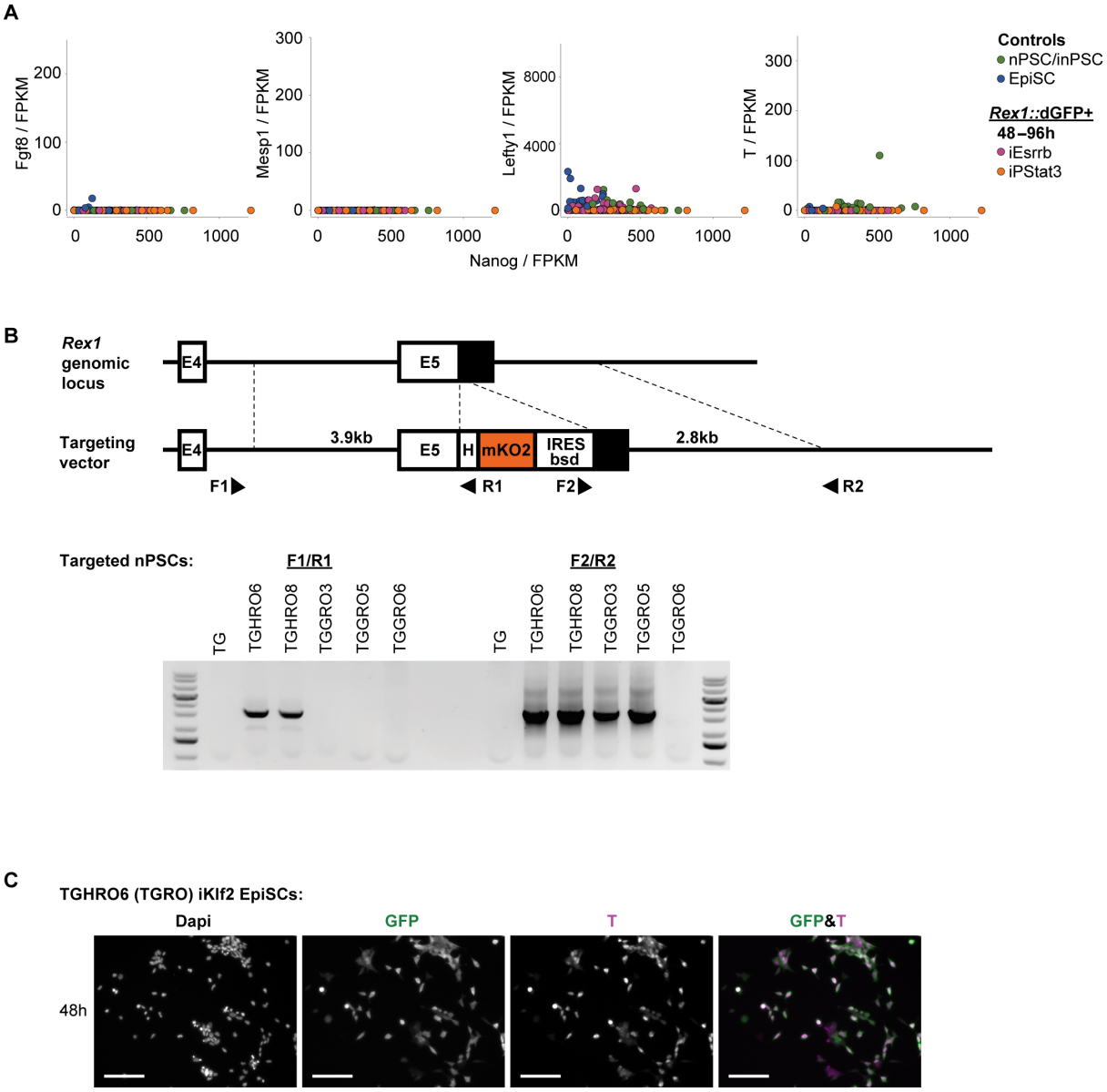

A
Fgf8 / FPKM
Mesp1 / FPKM
Lefty1 / FPKM
T / FPKM
Nanog / FPKM
Controls
nPSC/inPSC
EpiSC
Rex1::dGFP+
48–96h
iEsrrb
iPStat3
B
Rex1 genomic locus
Targeting vector
E4
E5
H
mKO2
IRES bsd
3.9kb
2.8kb
F1
R1
F2
R2
Targeted nPSCs:
F1/R1
F2/R2
TG
TGHRO6
TGHRO8
TGGRO3
TGGRO5
TGGRO6
C
TGHRO6 (TGRO) iKlf2 EpiSCs:
Dapi
GFP
T
GFP&T
48h

Figure S4

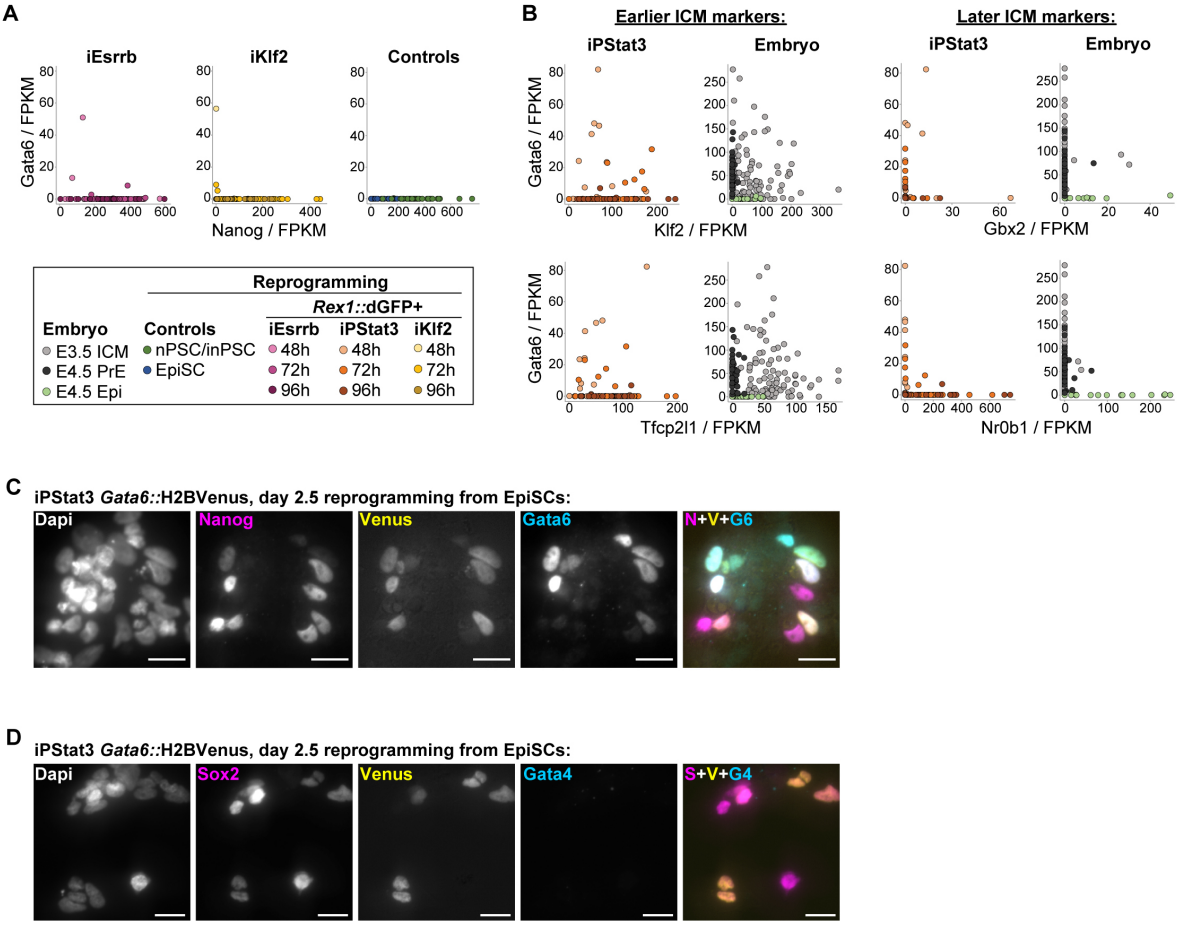

A
iEsrrb
iKlf2
Controls
Gata6 / FPKM
Nanog / FPKM
Reprogramming
Rex1::dGFP+
Embryo
Controls
iEsrrb
iPStat3
iKlf2
E3.5 ICM
E4.5 PrE
E4.5 Epi
nPSC/inPSC
EpiSC
48h
72h
96h
B
Earlier ICM markers:
Later ICM markers:
iPStat3
Embryo
Klf2 / FPKM
Gbx2 / FPKM
Tfcp2l1 / FPKM
Nr0b1 / FPKM
C
iPStat3 Gata6::H2BVenus, day 2.5 reprogramming from EpiSCs:
Dapi
Nanog
Venus
Gata6
N+V+G6
D
iPStat3 Gata6::H2BVenus, day 2.5 reprogramming from EpiSCs:
Dapi
Sox2
Venus
Gata4
S+V+G4

Figure S5

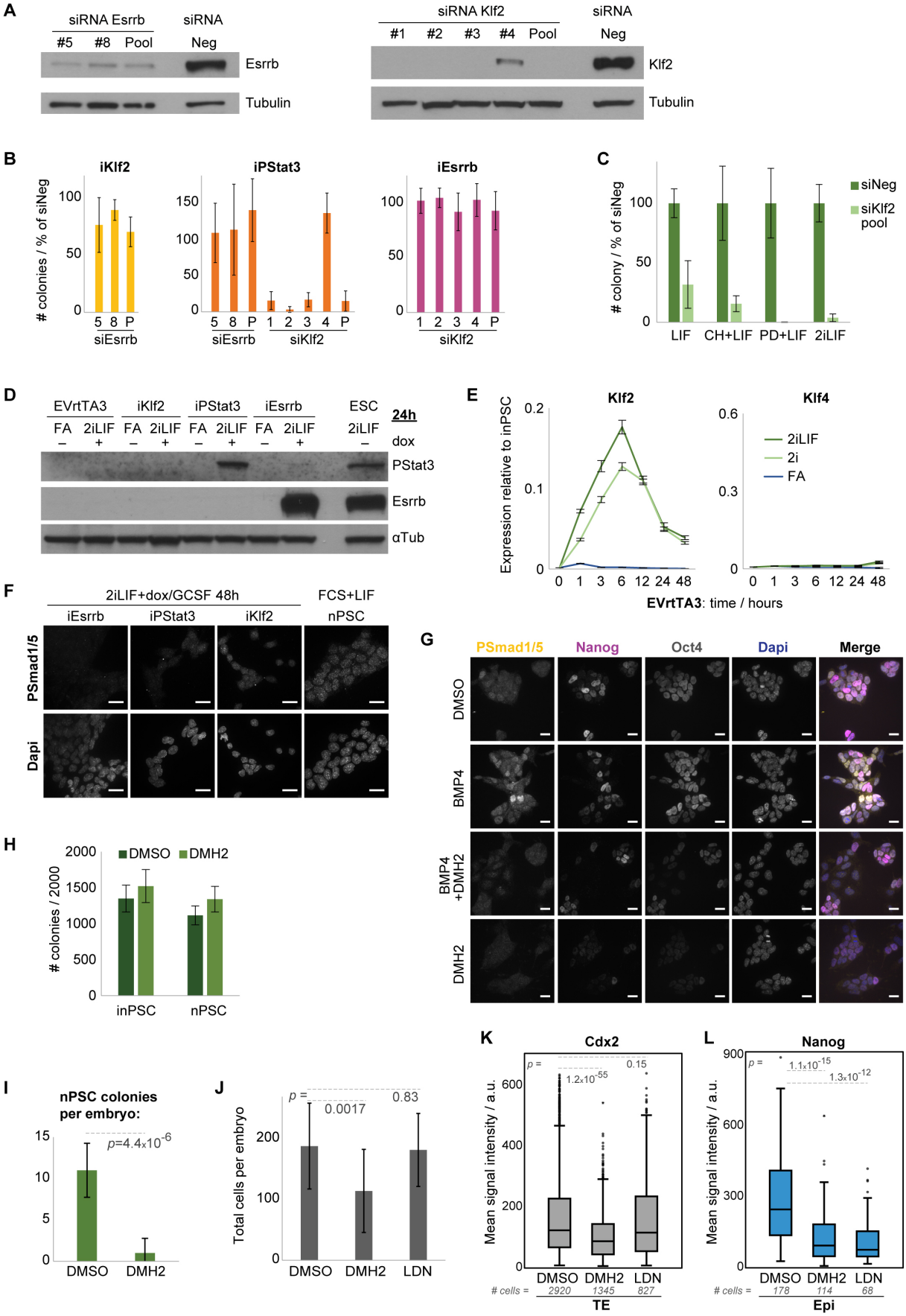

A
siRNA Esrrb
#5 #8 Pool
siRNA Neg
Esrrb
Tubulin
siRNA Klf2
#1 #2 #3 #4 Pool
siRNA Neg
Klf2
Tubulin
B
iKlf2
iPStat3
iEsrrb
colonies / % of siNeg
siEsrrb
siKlf2
C
colony / % of siNeg
siNeg
siKlf2 pool
LIF CH+LIF PD+LIF 2iLIF
D
EVrtTA3 iKlf2 iPStat3 iEsrrb ESC
FA 2iLIF FA 2iLIF FA 2iLIF FA 2iLIF 2iLIF
24h dox
PStat3
Esrrb
αTub
E
Klf2
Klf4
Expression relative to inPSC
2iLIF
2i
FA
EVrtTA3: time / hours
F
2iLIF+dox/GCSF 48h
FCS+LIF
iEsrrb iPStat3 iKlf2 nPSC
PSmad1/5
Dapi
G
PSmad1/5 Nanog Oct4 Dapi Merge
DMSO
BMP4
BMP4 +DMH2
DMH2
H
DMSO DMH2
colonies / 2000
inPSC nPSC
I
nPSC colonies per embryo:
p=4.4x10-6
DMSO DMH2
J
p =
0.0017
0.83
Total cells per embryo
DMSO DMH2 LDN
K
Cdx2
p =
1.2x10-55
0.15
Mean signal intensity / a.u.
DMSO DMH2 LDN
cells = 2920 1345 827
TE
L
Nanog
p =
1.1x10-15
1.3x10-12
Mean signal intensity / a.u.
DMSO DMH2 LDN
cells = 178 114 68
Epi

Figure S6

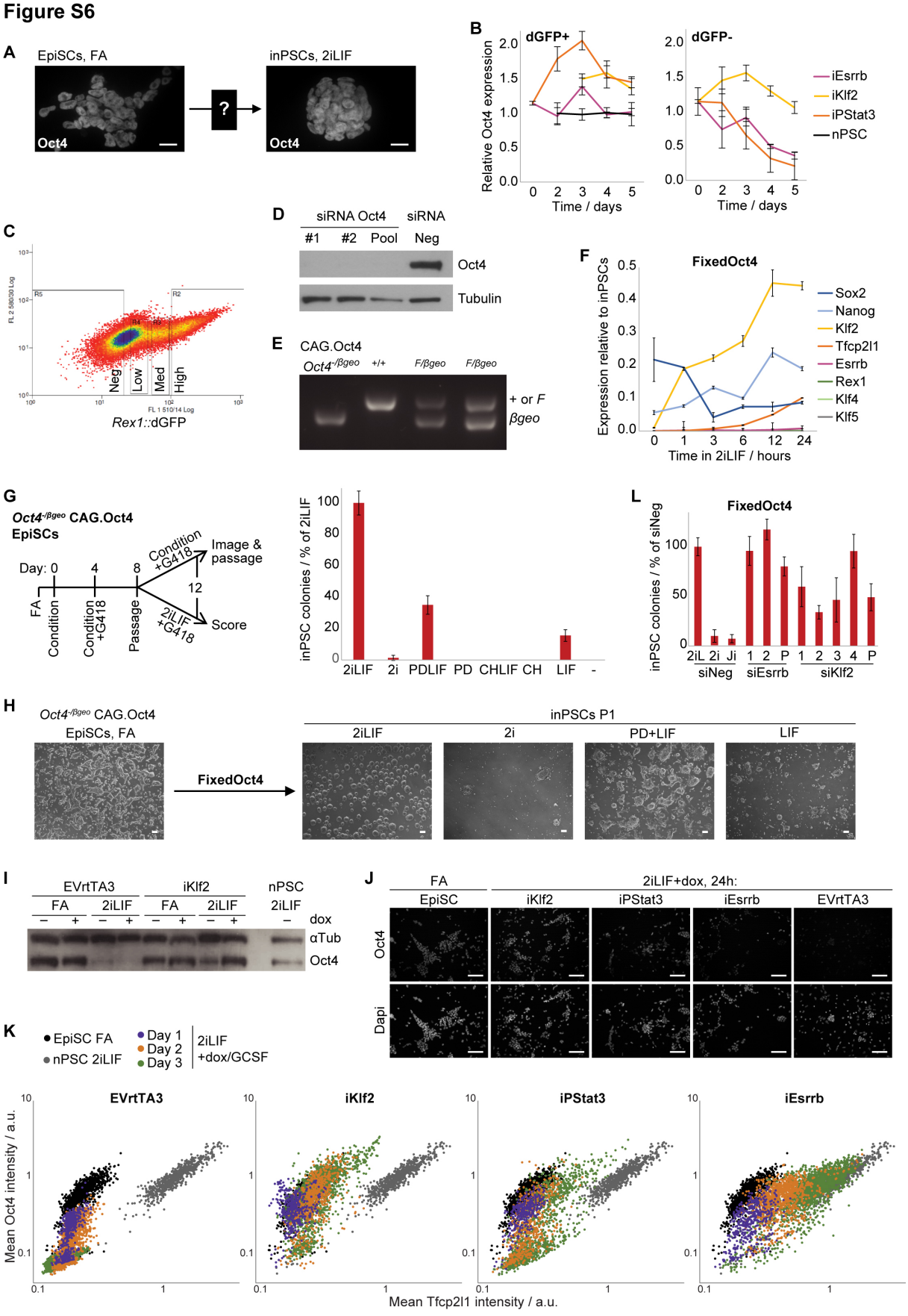
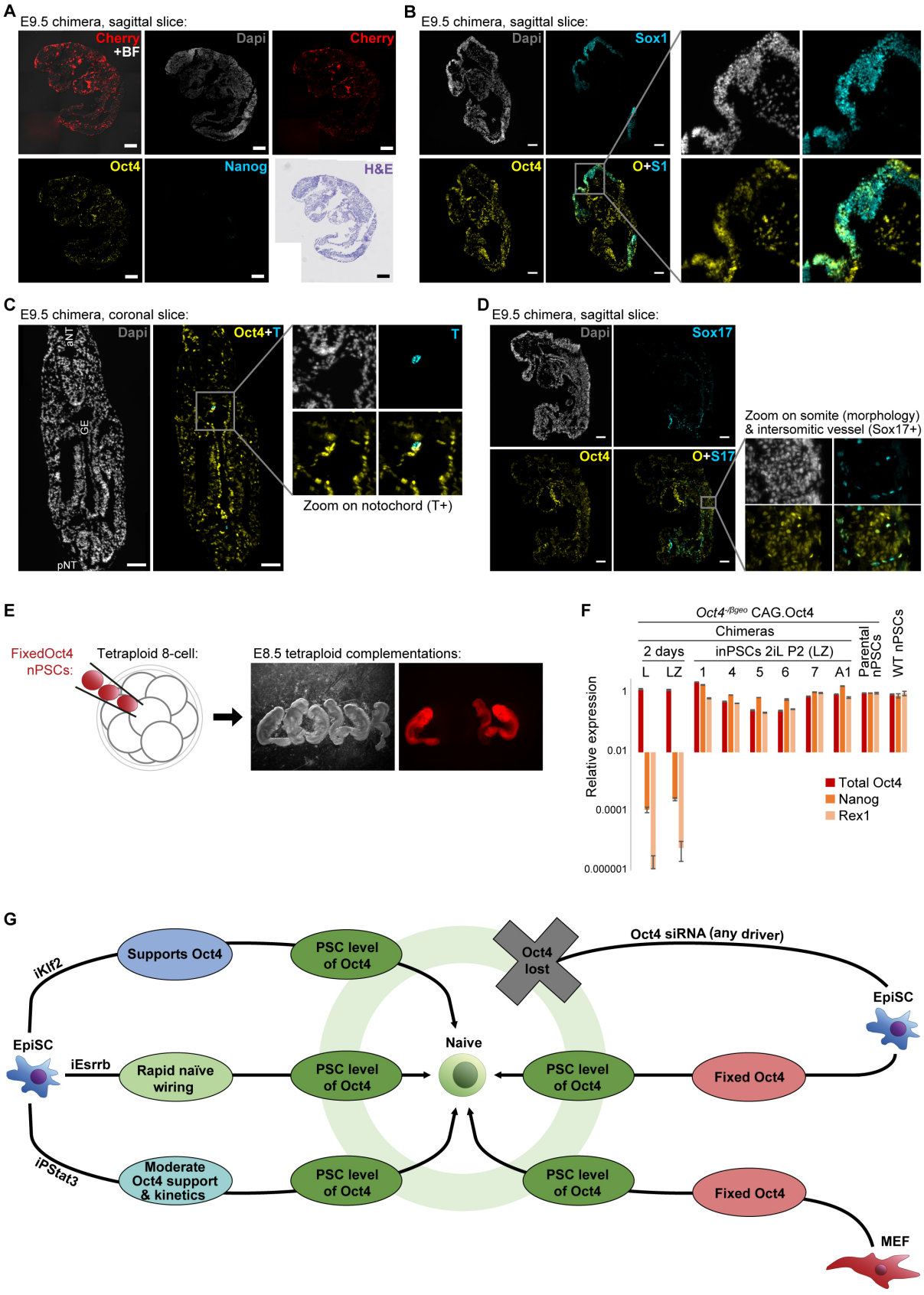

A
EpiSCs, FA
Oct4
?
inPSCs, 2iLIF
Oct4
B
dGFP+
dGFP-
Relative Oct4 expression
Time / days
iEsrrb
iKlf2
iPStat3
nPSC
C
Rex1::dGFP
Neg
Low
Med
High
D
siRNA Oct4
siRNA
#1
#2
Pool
Neg
Oct4
Tubulin
E
CAG.Oct4
Oct4-/βgeo
+/+
F/βgeo
F/βgeo
+ or F
βgeo
F
FixedOct4
Expression relative to inPSCs
Time in 2iLIF / hours
Sox2
Nanog
Klf2
Tfcp2l1
Esrrb
Rex1
Klf4
Klf5
G
Oct4-/βgeo CAG.Oct4
EpiSCs
Day: 0
4
8
12
FA
Condition
Condition +G418
Passage
Condition +G418
Image & passage
2iLIF +G418
Score
inPSC colonies / % of 2iLIF
2iLIF
2i
PDLIF
PD
CHLIF
CH
LIF
-
L
FixedOct4
inPSC colonies / % of siNeg
2iL
2i
Ji
1
2
P
1
2
3
4
P
siNeg
siEsrrb
siKlf2
H
Oct4-/βgeo CAG.Oct4
EpiSCs, FA
FixedOct4
inPSCs P1
2iLIF
2i
PD+LIF
LIF
I
EVrtTA3
iKlf2
nPSC
FA
2iLIF
FA
2iLIF
2iLIF
dox
αTub
Oct4
J
FA
2iLIF+dox, 24h:
EpiSC
iKlf2
iPStat3
iEsrrb
EVrtTA3
Oct4
Dapi
K
EpiSC FA
nPSC 2iLIF
Day 1
Day 2
Day 3
2iLIF
+dox/GCSF
EVrtTA3
iKlf2
iPStat3
iEsrrb
Mean Oct4 intensity / a.u.
Mean Tfcp2l1 intensity / a.u.

Figure S7


A
E9.5 chimera, sagittal slice:
Cherry +BF
Dapi
Cherry
Oct4
Nanog
H&E
B
E9.5 chimera, sagittal slice:
Dapi
Sox1
Oct4
O+S1
C
E9.5 chimera, coronal slice:
aNT
Dapi
GE
pNT
Oct4+T
T
Zoom on notochord (T+)
D
E9.5 chimera, sagittal slice:
Dapi
Sox17
Oct4
O+S17
Zoom on somite (morphology) & intersomitic vessel (Sox17+)
E
FixedOct4 nPSCs:
Tetraploid 8-cell:
E8.5 tetraploid complementations:
F
Oct4-/βgeo CAG.Oct4
Chimeras
2 days
inPSCs 2iL P2 (LZ)
L
LZ
1
4
5
6
7
A1
Parental nPSCs
WT nPSCs
Relative expression
1
0.01
0.0001
0.000001
Total Oct4
Nanog
Rex1
G
iKlf2
Supports Oct4
PSC level of Oct4
Oct4 lost
Oct4 siRNA (any driver)
EpiSC
EpiSC
iEsrrb
Rapid naïve wiring
PSC level of Oct4
Naive
PSC level of Oct4
Fixed Oct4
iPStat3
Moderate Oct4 support & kinetics
PSC level of Oct4
PSC level of Oct4
Fixed Oct4
MEF

Supplemental Tables

GO biological process	*p* value
supramolecular fiber organization (GO:0097435)	4.66E-08
cytoskeleton organization (GO:0007010)	3.89E-06
regulation of cellular component movement (GO:0051270)	4.52E-06
cellular component organization (GO:0016043)	5.00E-06
cellular component organization or biogenesis (GO:0071840)	2.56E-05
regulation of actin filament-based process (GO:0032970)	2.93E-05
regulation of localization (GO:0032879)	5.97E-05
regulation of biological quality (GO:0065008)	6.72E-05
regulation of multicellular organismal process (GO:0051239)	1.08E-04
regulation of supramolecular fiber organization (GO:1902903)	1.48E-04
regulation of cell motility (GO:2000145)	4.25E-04
actin cytoskeleton organization (GO:0030036)	4.72E-04
regulation of cell migration (GO:0030334)	5.01E-04
regulation of locomotion (GO:0040012)	6.52E-04
actin filament organization (GO:0007015)	7.41E-04
regulation of developmental process (GO:0050793)	1.04E-03
negative regulation of biological process (GO:0048519)	1.75E-03
anatomical structure development (GO:0048856)	1.99E-03
actin filament-based process (GO:0030029)	2.92E-03
negative regulation of cellular process (GO:0048523)	3.46E-03
regulation of system process (GO:0044057)	3.81E-03
protein localization to plasma membrane (GO:0072659)	4.32E-03
developmental process (GO:0032502)	5.19E-03
regulation of cell differentiation (GO:0045595)	5.80E-03
cellular process (GO:0009987)	8.64E-03
system development (GO:0048731)	9.78E-03
regulation of cytoskeleton organization (GO:0051493)	1.51E-02
regulation of actin cytoskeleton organization (GO:0032956)	1.84E-02
regulation of biological process (GO:0050789)	1.96E-02
regulation of cellular component organization (GO:0051128)	1.99E-02
animal organ development (GO:0048513)	2.90E-02
biological regulation (GO:0065007)	3.62E-02
regulation of cell population proliferation (GO:0042127)	4.60E-02

Table S1: Gene ontology enrichment for genes contributing to initial iKlf2 diversion. Gene ontology (GO) analysis for biological processes enriched in the gene list contributing to –PC1 dimension on the iKlf2 PCA plot (main Fig 2D left panel) (contribution score <–0.5). GO analysis was conducted using the PANTHER Overrepresentation Test with Bonferroni correction, and processes with $p < 0.05$ are presented in the table above.

Materials & Methods

Tissue culture

nPSCs and inPSCs were cultured in N2B27+2i+LIF (2iLIF). EpiSCs were cultured in N2B27+ XAV+FGF2+ActivinA (FA). N2B27 medium comprised 1:1 DMEM/F-12 and Neurobasal (Gibco), 2 mM L-glutamine (Gibco), 1x penicillin-streptomycin (Sigma), 0.1 mM 2-mercaptoethanol (Gibco), 1% B27 (Gibco) and 0.5% N2 (homemade). As required, N2B27 was supplemented with 20 ng/ml murine LIF (homemade), 3 μM CHIR99021 (Chiron; CH) and 1 μM PD0325901 (PD03; PD) (Stewart lab, Dresden), 12.5 ng/ml FGF2 and 20 ng/ml ActivinA (homemade), 6.25 μg/ml XAV 939 (Tocris), 3 μM DMH2 (Tocris), or 0.6 μM LDN193189 (Sigma). For nPSCs and inPSCs, tissue-culture flasks were coated with 0.15% gelatin (Sigma) in PBS (Sigma) and incubated in 7% CO_2. For EpiSC culture and reprogramming experiments, tissue-culture flasks were coated with 10 μg/ml fibronectin (Millipore) in PBS (Sigma) and incubated in 7% CO_2 and 5% O_2. nPSCs, inPSCs and EpiSCs were dissociated with accutase (Millipore) during passaging. For optimal performance of EpiSCs, lines were maintained by plating 25000 cells/cm^2 every other day (usually 1:6 split ratio) following gentle accutase treatment for less than 3 minutes at room temperature. For Figures S5F–G and 6D, nPSCs were cultured in FCS+LIF medium containing GMEM (Sigma), 10% fetal calf serum (FCS) (Gibco), 1x non-essential amino acids (Gibco), 1 mM sodium pyruvate (Sigma), 2 mM L-glutamine (Gibco), 1x penicillin-streptomycin (Sigma), 0.1 mM 2-mercaptoethanol (Gibco), 20 ng/ml murine LIF (Hyvonen lab, Cambridge), and 10 ng/ml BMP4 (Miltenyi Biotec) was supplemented as indicated.

*Rex1::*dGFP EpiSCs

Rex1$^{dGFP.IRES.bsd/dGFP.IRES.bsd}$ homozygous 129 studs (Kalkan et al., 2017) were crossed with wild-type 129 females and heterozygous *Rex1*$^{+/dGFP.IRES.bsd}$ EpiSCs (referred to as *Rex1::*dGFP reporter) were derived from resultant E6.5 embryos. Epiblasts were manually dissected from extra-embryonic tissues and plated on fibronectin-coated plates in FA medium. After 5–7 days of culture, regions of the explant exhibiting EpiSC morphology were manually passaged to a fresh plate. Subsequent passages were performed using accutase (Millipore).

Cell transfection

For transgene integration transfections, 1 μg PiggyBac (PB) vectors of interest, 0.5 μg PBase expression vector (*CAG.PBase*) and 10 μl Lipofectamine-2000 (Invitrogen) were incubated for 20 min in 500 ml DMEM (Gibco), then applied to 500,000 cells/6well in 3 ml medium for 18 hours. Selection was applied to transfectants for at least 5 passages prior to use: 50 μg/ml hygromycin-B (Life Technologies) for *PB.TetO.GOI.PGK.hph* or *PB.CAG.GY118F.PGK.hph*, and 0.33 μg/ml puromycin (ThermoFisher) for *PB.CAG.rtTA3.PGK.pac*. siRNA transfections were performed using RNAiMAX transfection reagent (Invitrogen) and FlexiTube siRNAs against Oct4, Klf2, Esrrb, or AllStars Negative Control (Qiagen) according to the manufacturers' instructions.

EpiSC reprogramming

EpiSCs were plated in FA without selection at a density of 2000/24well or equivalent. For siRNA experiments, 10000/24well or equivalent was used instead to compensate for transfection toxicity. The following day, reprogramming was induced by medium change to 2iLIF or subset components thereof as indicated. GY118F transgenic receptor was stimulated with 30 ng/ml human GCSF (Peprotech), whereas expression of other transgenes was induced with 1 μg/ml doxycycline (MP Biomedicals). After 4 days, transgene induction was withdrawn and 20 μg/ml blasticidin (Gibco) was applied to select for *Rex1::*dGFP.IRES.bsd activity. On day 8, 4x images were acquired using CellSens software and an X-51 Olympus microscope system with motorized stage and camera. inPSC colonies with active *Rex1* reporter were counted manually. *Rex1* reporter activity confers both dGFP expression and blasticidin resistance, and is a well-characterised naïve marker (Kalkan et al., 2017). We confirmed that dGFP+ colonies are also Oct4+Tfcp2l1+ by immunostaining (data not shown). Only 4 days of

transgene induction is a stringent test of driver efficacy; we note that more colonies emerged when induced for longer, including for weaker drivers iNanog and iTfcp2l1 (data not shown). No colonies ever emerged from any EVrtTA3 reprogramming experiments, confirming that our lines represent “late-stage” EpiSCs (Han et al., 2010). Where indicated, reprogramming experiments were treated with 3 μM DMH2 (Tocris), 0.6 μM LDN (Sigma), or 1 μM Jak inhibitor (Merck-Millipore) from days 0–4. Unless stated otherwise, reprogramming data presented are the mean of 3 biological replicates.

FixedOct4 EpiSCs
FixedOct4 EpiSCs were generated from *Oct4*$^{-/\beta geo}$ CAG.Oct4^{wt}.2A.mCherry nPSCs (Radzisheuskaya et al., 2013) by differentiation in FA for 10 passages. *Oct4*$^{F/\beta geo}$ CAG.EmptyVector EpiSCs were generated as a control from the same parental line. Reprogramming was conducted in N2B27+2iLIF as above. 200 μg/ml G418 (Invitrogen) was applied to select for endogenous *Oct4* promoter activity after 4 days of reprogramming.

***Gata6::*H2BVenus EpiSCs**
Gata6$^{+/H2BVenus}$ nPSCs (Freyer et al., 2015) were kindly shared by Christian Schröter. By differentiation for 10 passages in FA, we obtained *Gata6::*H2BVenus reporter EpiSCs.

***T::*GFP *Rex1::*mKO2 EpiSCs**
T$^{+/GFP}$ nPSCs (Fehling et al., 2003) were kindly shared by Gordon Keller. Rex1-mKO2 fusion cassette was constructed by replacing the dGFP cassette of the Rex1-dGFP targeting vector (Kalkan et al., 2017) (Figure S3B). We linearized the vector with BspH1, then electroporated it into *T*$^{+/GFP}$ nPSCs using Gene Pulser (BioRad) at 230V, 500μF. Correct targeting results in a Rex1-mKO2 fusion protein and confers blasticidin resistance when *Rex1* is expressed. nPSCs were selected with 10 μg/ml blasticidin, then clones were genotyped by PCR. PCR primers for 5’ side are TCGTGTGACTCTGCATCTGT and CTGCCTCTTTAGCTGCGG, and for 3’ side are ATTCGTGAATTGCTGCCCTC and GAGGCAGAGGAACAGGACTT. Correctly targeted nPSC clone TGHRO6 (subsequently referred to as TGRO) was differentiated to EpiSCs by 10 passages in FA, resulting in *T::*GFP *Rex1::*mKO2 double-reporter EpiSCs.

Live imaging
Live imaging was performed using IncuCyte system, with phase and H2BVenus images taken every 60 min for *Gata6* reporter, or phase, GFP and mKO2 images taken every 45 min for *T/Rex1* double-reporter. For *Gata6* reporter, the endpoint was fixed, stained for Tfcp2l1 (AF594) and Oct4 (AF647), then re-imaged with the same positional registration for AF594. Co-expression of Tfcp2l1+Oct4+ in endpoint inPSCs was confirmed on a separate microscope capable of detecting AF647 as well (data not shown).

Microinjection to generate chimeras
Chimeras were generated from strain 129 (agouti) male inPSCs by standard microinjection methodology using host blastocysts of strain C57BL/6 (black), followed by gestation in pseudo-pregnant recipient females. Germline-competence of male chimeras was tested by crossing them to C57BL/6 (black) females and checking for agouti pups. For Figure 4G–I, host embryos were injected at the 8-cell stage, cultured in Blast medium (Origio) until blastocysts formed, then cultured to the late blastocyst stage in N2B27. For Figure S7E, tetraploid host embryos were generated by cell fusion at the 2-cell stage, cultured to the 8-cell stage in Cleav medium (Origio), injected then transferred to pseudo-pregnant recipients for gestation. All animal work was performed in accordance with Home Office guidelines and regulations at the University of Cambridge, UK.

Reprogramming from FixedOct4 E9.5 chimeras
E9.5 chimeras were generated by blastocyst injection of *Oct4*$^{-/\beta geo}$ CAG.Oct4^{wt}.2A.mCherry nPSCs (Radzisheuskaya et al., 2013). The tail portions of resulting E9.5 embryos were re-

moved to strictly avoid germ cell contamination in the cultures. The anterior portion was dissociated manually, then subdivided into quarters (2x LIFaza, 2x 2iLIF or LIF only as indicated). Aza = 1 μM 5-Azacytidine (Sigma). After 6 days, LIF or LIFaza was exchanged for 2iLIF, then on day 10 G418 was applied to all cultures (200 μg/ml). Chimeric allantois samples were dissociated manually, a portion taken for expression analysis, and the remainder plated as positive control (germ-cell containing).

Derivation and reprogramming of FixedOct4 MEFs

E12.5 chimeras were generated by blastocyst injection of *Oct4*$^{-/\beta geo}$ CAG.Oct4^{wt}.2A.mCherry nPSCs (Radzisheuskaya et al., 2013). The heads and all internal organs were removed, taking particular care to fully remove the genital ridges. Carcasses were dissociated in trypsin (Life Technologies), then cultured in FCS medium. Hygromycin was applied to select for CAG.Oct4^{wt}.2A.mCherry transgene. MEFs were passaged using trypsin, and used from passage 3–5 for reprogramming assays. For reprogramming, MEFs were plated at 5000/24well in FCS medium on 0.15% gelatin. The following day, medium was changed to N2B27+LIF±IL6&IL6R±FGF2±CH±PD as indicated for the first week (20 ng/ml LIF; 50 ng/ml IL6 and 10 ng/ml soluble IL6R; 12.5 ng/ml FGF2; 3 μM CH; 1 μM PD). On day 7, all were swapped to N2B27+2iLIF, then on day 14 G418 was added (200 μg/ml) to select for inPSC colonies. We would like to highlight that all medium was N2B27-based, i.e. MEF reprogramming occurred in the absence of serum, KSR/ascorbic acid, or any small molecule epigenetic modulators.

BMP inhibitor treatment of embryos

Wild-type 129 mice were crossed and embryos flushed from oviducts at 2.75 dpc using M2 medium (Millipore). Embryos were subsequently incubated in Blast medium (Origio) and periodically inspected. At cavitation onset, embryos were randomly divided into Blast medium supplemented either with 3 μM DMH2 (Tocris), 0.3 μM LDN (Sigma) or 1:1000 DMSO. Once blastocysts had fully formed, they were transferred to N2B27 medium continuing DMH2/LDN/DMSO treatment as before. At the late blastocyst stage, embryos were fixed and immunostaining was performed. Embryos were permeabilized in 0.25% Triton X-100 (Sigma) in PBS for 30 min, then blocked in 3% donkey serum (Sigma), 0.1% BSA (Sigma) and 0.01% Tween-20 (Sigma) for 30 min at room temperature. Embryos were incubated overnight at 4°C in blocking buffer with the following primary antibodies: Cdx2 (1:500, mouse mAb, BioGenex); Gata4 (1:300, goat pAb, Santa Cruz); Nanog (1:300, rat mAb, eBioscience); Oct4 (1:300, rabbit mAb, Cell Signaling). The following day, washes were performed in blocking buffer. AlexaFluor secondary antibodies (Life Technologies) were used against the appropriate species at 1:1000 in blocking buffer. Embryos were gradually acclimatised then mounted in Fluoromount-G (Southern Biotech) and images were taken with a Zeiss 710 LSM confocal microscope. Presented images are maximum intensity projections of Z-stack slices processed with ImageJ. Staining quantification was carried out with Imaris: nuclei were identified in the DAPI channel and the fluorescence of each other channel recorded. Cells were assigned to each lineage based on position and marker staining. We note that the phenotype was highly time-sensitive: addition of BMP inhibitor at the 8-cell stage caused developmental arrest, consistent with Reyes de Mochel et al., 2015 but precluding fair assessment of whether the naïve lineage specifically is compromised. At the 8-cell stage, the trophectoderm (TE) vs inner cell mass (ICM) decision has not yet been made. Conversely, application of BMP inhibitor to the mid blastocyst did not disrupt the naïve epiblast (data not shown). By this point, the epiblast vs primitive endoderm (PrE) bifurcation is already underway. In contrast, precisely timed inhibitor addition at cavitation onset in the late morula falls between these two developmental lineage bifurcations, and thus permits assessment of the role of BMP signalling in naïve epiblast establishment despite the multitude of BMP signalling roles during early development.

Quantitative nPSC derivation

Following BMP inhibitor treatment from cavitation onset as above, quantitative nPSC-derivation was performed from late blastocysts as previously described (Nichols et al., 2009).

Briefly, immunosurgery was performed to remove the TE, then the ICM (comprising epiblast+PrE) was dissociated to single cells using accutase. 10 single cells were manually transferred to each 96well and cultured in feeder-free N2B27+2iLIF conditions on gelatin, without any further DMSO/DMH2 treatment so that we could assess whether the epiblast of the embryo was already affected. The number of nPSC colonies was scored after 6 days, and nPSC identity confirmed by RT-qPCR (data not shown).

Flow cytometry

Flow cytometry was performed using a BD LSRFortessa analyser with subsequent data analysis using FlowJo software. Cell sorting was performed using a MoFlo Legacy Cell Sorter (Beckman) or an S3 Cell Sorter (BioRad). dGFP was excited using a 488 nm laser and detected using a 530/30 filter. *Rex1::dGFP* EpiSCs and nPSCs were used to determine negative and positive dGFP gates respectively. After sorting of reprogramming intermediates, number of inPSC colonies are quantified relative to the number of nPSC colonies, because replating of sorted nPSCs provides a control for cell death due to the stress of sorting. nPSCs already stably occupy the destination naïve pluripotent identity, and are thus the appropriate functional control. When we replated *Rex1::*dGFP+ reprogramming intermediates for clonogenicity assay, we later applied blasticidin as an additional control to prove that the *Rex1* promoter was active in scored inPSC colonies: *Rex1* promoter drives both dGFP expression and blasticidin resistance.

Immunohistochemistry

Cultured cells were fixed with 4% paraformaldehyde (Sigma) in PBS for 10 min at room temperature. E9.5 embryos were fixed with 4% paraformaldehyde (Sigma) in PBS for 4 hours at 4°C, gradually adjusted to 20% sucrose over 2 days, mounted in O.C.T. (TissueTek), snap frozen on liquid nitrogen, cryosectioned (8 µm), stored at -80°C, then rehydrated in PBS. For both cultured cells and embryo cryosections, permeabilization was performed in 0.4% Triton X-100 (Sigma) in PBS, then samples were blocked in 5% donkey serum (Sigma) and 0.1% Triton X-100 in PBS. Samples were incubated overnight at 4°C in blocking buffer with the following primary antibodies: Cardiac troponin (1:300, mouse mAb, Abcam); Esrrb (1:300, mouse mAb, Perseus Proteomics); FoxA2 (1:300, goat pAb, R&D); Gata4 (1:300, goat pAb, Santa Cruz); Gata6 (1:300, goat Santa Cruz); GFP (1:300, rat mAb, Nacalcai); Klf2 (kind gift from Hitoshi Niwa (1:300, mouse mAb, Yamane et al., 2018)); Klf4 (1:300, goat pAb, R&D); Nanog (1:300, rat mAb, eBioscience); Oct4 (1:300, rabbit mAb, Cell Signaling); Oct4 (1:100, mouse mAb, Santa Cruz); PSmad1/5 (1:100, rabbit mAb, Cell Signaling); P-Y705-Stat3 (1:300, rabbit mAb, Cell Signaling); Sox1 (1:300, goat pAb, R&D); Sox2 (1:300, rat mAb, eBioscience); Sox17 (1:300, goat pAb, R&D); T (1:300, goat pAb, R&D); Tfcp2l1 (1:300, goat pAb, R&D). The next day, washes were performed with 0.1% Triton X-100 in PBS, and samples incubated with DAPI (ThermoFisher) and AlexaFluor secondary antibodies against the appropriate species at 1:1000 (Life Technologies). For PSmad1/5 and PStat3 stainings, TBS was used at all steps instead of PBS. Samples were mounted in Fluoromount-G (Southern Biotech) and imaged using Leica DMI6000, Nikon Eclipse Ti Spinning Disk confocal or Zeiss ApoTome microscope. Staining quantification was carried out with CellProfiler: nuclei were identified in the DAPI channel and the fluorescence of each other channel recorded. Confocal images are presented as maximum intensity projections of Z-stack slices processed with ImageJ. Embryo sections were imaged using the Zeiss ApoTome microscope at 20x then tiled. After imaging, H&E histological staining was performed on the same sections according to standard methodologies. These sections were then re-imaged in the same pipeline.

Western blotting

Cells were lysed in RIPA buffer (Sigma) containing Complete-ULTRA protease-inhibitor and PhoStop phosphatase-inhibitor cocktails (Roche), and sonicated with Bioruptor200 (Diagenode) at high frequency, alternating 30 s on/off for 3 min. SDS-PAGE electrophoresis was performed using Bolt 10% Bis-Tris Plus gels (ThermoFisher) in a Novex MiniCell (ThermoFisher). Protein transfer was performed using the semi-dry iBlot2 system (ThermoFisher)

and iBlot Transfer Stacks (ThermoFisher). The following primary antibodies were used: Esrrb (1:1000, mouse mAb, R&D); Klf2 (kind gifts from Huck-Hui Ng (1:500, rabbit serum, Yeo et al., 2014) and Hitoshi Niwa (1:1000, mouse mAb, Yamane et al., 2018)); Oct4 (1:1000, rabbit mAb, Cell Signaling); P-Y705-Stat3 (1:1000, rabbit mAb, Cell Signaling); αTubulin (1:10000, mouse mAb, Abcam). Detection was achieved using HRP-linked secondary antibodies at 1:10000 against the appropriate species (GE Healthcare) and ECL Plus Western Blotting Detection System (GE Healthcare).

Genotyping

Genotyping to distinguish between *Oct4*$^{-/\beta geo}$, *Oct4*$^{F/\beta geo}$ and *Oct4*$^{+/+}$ cells was conducted using Taq DNA Polymerase (Qiagen) according to manufacturer's instructions and the following thermocycler program: 95°C 3 min; 30x 94°C 15 s, 60°C 30 s, 72°C 60 s; 72°C 10 min. Primer GAGCTTATGATCTGATGTCCATCTCTGTGC binds in the *Oct4* final intron, which is present in both wild-type and Flox (F) alleles. Primer GGGCTGACCGCTTCCTCGTGCTTTACG binds in the *βgeo* allele. Primer GCCTTCCTCTATAGGTTGGGCTCCAACC binds 3' downstream of Oct4 and is common to all alleles.

Plasmids

PiggyBac expression vector was modified to contain TetO rather than CAG promoter. Genes of interest (Esrrb, Klf2, Klf4, Klf5, Nanog, Tfcp2l1) were cloned into the resultant vector using Gateway technology (ThermoFisher) according to manufacturer's instructions, producing inducible expression vectors in the form *PB.TetO.GOI.PGK.hph*. GY118F and rtTA3 expression vectors were generated in the same manner, but retaining CAG instead of TetO promoters and thus yielding *PB.CAG.GY118F.PGK.hph* and *PB.CAG.rtTA3.PGK.pac* respectively. *PB.CAG.GY118F.PGK.bsd* was used in the *Gata6::*H2BVenus reporter line.

RT-qPCR

Total RNA was extracted using RNeasy kits, according to manufacturer's spin protocol, including on-column DNaseI digest (Qiagen). cDNA was produced from 1 µg RNA using SuperscriptIII VILO cDNA synthesis kit, following the recommended protocol including RNaseH treatment (Invitrogen). RT-qPCR reactions were performed using StepOnePlus Real Time PCR System with recommended thermocycler settings (Applied Biosystems) and TaqMan Fast Universal PCR Master Mix (Applied Biosystems). Gene expression relative to Gapdh in each well was determined using FAM-labelled TaqMan assay probe together with VIC-labelled Gapdh probe (Applied Biosystems).

Gene	Probe ID
Gapdh	4352339E
Esrrb	Mm00442411_m1
Fgf5	Mm00438918_m1
Gata6	Mm00802636_m1
Klf2	Mm01244979_g1
Klf4	Mm00516104_m1
Klf5	Mm00456521_m1
Nanog	Mm02384862_g1
Nr0b1	Mm00431729_m1
Nr5a2	Mm00446088_m1
Pou5f1	Mm00658129_gH
Sox2	Mm03053810_s1
T (Brachyury)	Mm01318252_m1
Tfcp2l1	Mm00470119_m1
Zfp42	Mm03053975_g1

Bulk RNAseq library preparation
Sequencing libraries were prepared according to the SmartSeq2 protocol (Picelli et al., 2014) with the following amendments: purified RNA was used diluted to 5 ng/µl; ERCC spike-ins (Invitrogen) were added at 1 µl of 1:10000 dilution per 5 ng; 13 cycles of amplification were used to obtain cDNA (rather than 21 cycles used for single-cells). Nextera XT reactions were scaled-down by half, using 0.4ng cDNA input per reaction. Pooled libraries were sequenced on the Illumina HiSeq 4000 (paired-end 150bp reads).

scRNAseq library preparation
Single cells were index-sorted individually by FACS (BD Influx 5) into wells of a 96-well PCR plate containing lysis buffer. scRNAseq was performed as previously described (Nestorowa et al., 2016; Picelli et al., 2014; Wilson et al., 2015) for a total of 1152 single cells. The Illumina Nextera XT DNA kit was used to prepare libraries. Pooled libraries were sequenced on the Illumina HiSeq 4000 (single-end 125bp reads). Samples from all cell lines were included in each sequencing lane, to control for technical lane effects. We did not detect a significant batch effect.

RNAseq alignment
GENCODE M12 mouse gene annotation from Ensembl release 87 was used for read alignment (Yates et al., 2016) and splice junction donor/acceptor overlap settings were tailored to the read length of each dataset. Alignments to gene loci were quantified with HTSeq-count (Anders et al., 2015) based on annotation from Ensembl release 87. Sequencing libraries with fewer than 500,000 mapped reads were excluded from subsequent analyses. Read distribution bias across gene bodies was computed as the ratio between the total reads spanning the 50^{th} to the 100^{th} percentile of gene length, and those between the first and 49^{th}. Samples with ratio >2 were not considered further.

Published embryo scRNAseq datasets
Sequencing data of single-cell mouse embryo profiling studies SRP110669 (Mohammed et al., 2017; E3.5, E4.5, E6.5) and SRP020490 (Deng et al., 2014; compacted morula) were downloaded from the European Nucleotide Archive (Toribio et al., 2017) and aligned as above.

Transcriptome analysis
Principal component and cluster analyses were performed based on $\log_2$ FPKM values and were computed with the Bioconductor packages *DESeq2* (Love et al., 2014), *Sincell* (Juliá et al., 2015) or *FactoMineR* (Lê et al., 2008) in addition to custom scripts. Default parameters were used unless otherwise indicated. For global analyses, genes that registered zero counts in all single-cell samples were omitted. Euclidean distance and average agglomeration methods were used for cluster analyses unless otherwise indicated. t-SNE analysis was computed using Rtsne R package (Krijthe, 2015) with default parameters. *k*-means hard clustering was performed using the Mfuzz R package (Futschik and Carlisle, 2005) and the optimal number of *k* clusters were selected using the elbow method.

Selection of high-variability genes
Genes exhibiting the greatest expression variability (and thus contributing substantial discriminatory power) were identified by fitting a non-linear regression curve between average $\log_2$ FPKM and the square of coefficient of variation. Indicated specific thresholds were applied along the *x*-axis (average $\log_2$ FPKM) and *y*-axis (CV^2) to identify the most variable genes.

Quadratic programming
Fractional identity of single cells was computed using R package DeconRNASeq (Gong and Szustakowski, 2013). This package utilises quadratic programming to estimate the proportion of distinctive cell types. The average expression of bulk RNAseq samples (EpiSCs or inPSCs, Fig S2C) or the average expression of scRNAseq samples (embryo stages, Fig 4B) were used as signature datasets. The fraction of identity between single cells and the signature datasets

was computed using the whole transcriptome. Reprogramming pseudotimes for single cells were assigned by ordering cells based on their fraction of similarity to EpiSCs (origin) and inPSCs (destination) (Fig S2C), as previously described (Treutlein et al., 2016).

LOESS regression
Smooth curve local regression (LOESS) lines were fitted to scatter plots of $\log_2$ FPKM vs pseudotime using the R stats package (R Core Team, 2016). Smoothness parameter of 1/3 and 2 degrees of local polynomial were used for curve fitting.

Differential gene expression
Differential expression (DE) analysis was performed on each sample set relative to start EpiSCs, using the R package “scde" (Kharchenko et al., 2014) which fits individual error models for assessment of differential expression between groups of single cells.

Statistical analyses
Where appropriate, two-sided t tests were performed. *p*-values are indicated on the Figures.

Supplemental References

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science 343, 193–196.

Fehling, H.J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G., and Kouskoff, V. (2003). Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. Development 130, 4217–4227.

Freyer, L., Schröter, C., Saiz, N., Schrode, N., Nowotschin, S., Martinez-Arias, A., and Hadjantonakis, A.-K. (2015). A loss-of-function and H2B-Venus transcriptional reporter allele for Gata6 in mice. BMC Dev. Biol. 15, 38.

Futschik, M.E., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. J. Bioinform. Comput. Biol. 3, 965–988.

Gong, T., and Szustakowski, J.D. (2013). DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. Bioinformatics 29, 1083–1085.

Han, D.W., Tapia, N., Joo, J.Y., Greber, B., Araúzo-Bravo, M.J., Bernemann, C., Ko, K., et al. (2010). Epiblast Stem Cell Subpopulations Represent Mouse Embryos of Distinct Pregastrulation Stages. Cell 143, 617–627.

Juliá, M., Telenti, A., and Rausell, A. (2015). Sincell: an R/Bioconductor package for statistical assessment of cell-state hierarchies from single-cell RNA-seq. Bioinformatics 31, 3380–3382.

Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H.J., Nett, I., Marks, H., Walker, R., et al. (2017). Tracking the embryonic stem cell transition from ground state pluripotency. Development 144, 1221–1234.

Kharchenko, P. V, Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat. Methods 11, 740–742.

Krijthe, J.H. (2015). Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation.

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. J. Stat. Softw. 25.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. 20, 1215–1228.

Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Göttgens, B. (2016). A single cell resolution map of mouse haematopoietic stem and progenitor cell differentiation. Blood.

Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development 136, 3215–3222.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Radzisheuskaya, A., Chia Gle, B., dos Santos, R.L., Theunissen, T.W., Castro, L.F., Nichols, J., and Silva, J.C. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nat Cell Biol 15, 579–590.

Reyes de Mochel, N.S., Luong, M., Chiang, M., Javier, A.L., Luu, E., Toshihiko, F., MacGregor, G.R., et al. (2015). BMP signaling is required for cell cleavage in preimplantation-mouse embryos. Dev. Biol. 397, 45–55.

Toribio, A.L., Alako, B., Amid, C., Cerdeño-Tarrága, A., Clarke, L., Cleland, I., Fairley, S., Gibson, R., Goodgame, N., Ten Hoopen, P., et al. (2017). European Nucleotide Archive in 2016. Nucleic Acids Res. 45, D32–D36.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., et al. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature 534, 391–395.

Wilson, N.K., Kent, D.G., Buettner, F., Shehata, M., Macaulay, I.C., Calero-Nieto, F.J., Sánchez Castillo, M., Oedekoven, C.A., Diamanti, E., Schulte, R., et al. (2015). Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell 16, 712–724.

Yamane, M., Ohtsuka, S., Matsuura, K., Nakamura, A., and Niwa, H. (2018). Overlapping function of klf family targets multiple transcription factors to maintain naïve pluripotency of ES cells. Development dev.162404.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic Acids Res. 44.

Yeo, J.-C., Jiang, J., Tan, Z.-Y., Yim, G.-R., Ng, J.-H., Göke, J., Kraus, P., Liang, H., et al. (2014). Klf2 Is an Essential Factor that Sustains Ground State Pluripotency. Cell Stem Cell 14, 864–872.

Current Biology 24, 340–346, February 3, 2014 ©2014 The Authors. Open access under CC BY license. http://dx.doi.org/10.1016/j.cub.2013.12.040

Report

NANOG Amplifies STAT3 Activation and They Synergistically Induce the Naive Pluripotent Program

Hannah T. Stuart,[1,2] Anouk L. van Oosten,[1,2] Aliaksandra Radzisheuskaya,[1] Graziano Martello,[1] Anzy Miller,[1,2] Sabine Dietmann,[1] Jennifer Nichols,[1,3] and José C.R. Silva[1,2,*]

[1]Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK

[2]Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK

[3]Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK

Summary

Reprogramming of a differentiated cell back to a naive pluripotent identity is thought to occur by several independent mechanisms. Two such mechanisms include NANOG and activated STAT3 (pSTAT3), known master regulators of naive pluripotency acquisition [1–5]. Here, we investigated the relationship between NANOG and pSTAT3 during the establishment and maintenance of naive pluripotency. Surprisingly, we found that NANOG enhances LIF signal transduction, resulting in elevated pSTAT3. This is mediated, at least in part, by suppression of the expression of the LIF/STAT3 negative regulator SOCS3. We also discovered NANOG to be limiting for the expression of KLF4, a canonical "Yamanaka" reprogramming factor [6] and key pSTAT3 target [2, 7, 8]. KLF4 expression resulted from the codependent and synergistic action of NANOG and pSTAT3 in embryonic stem cells and during initiation of reprogramming. Additionally, within 48 hr, the combined actions of NANOG and pSTAT3 in a reprogramming context resulted in reactivation of genes associated with naive pluripotency. Importantly, we show that NANOG can be bypassed during reprogramming by exogenous provision of its downstream effectors, namely pSTAT3 elevation and KLF4 expression. In conclusion, we propose that mechanisms of reprogramming are linked, rather than independent, and are centered on a small number of genes, including NANOG.

Results

NANOG Amplifies STAT3 Activation

We first investigated the effect of NANOG on STAT3 activation in embryonic stem cells (ESCs), since the pluripotency network is established and functional in this cellular context. Although NANOG and STAT3 are essential for embryonic naive pluripotency establishment [4, 9], they promote but are not required for in vitro ESC maintenance [10–14]. This permits the study of *Nanog*$^{-/-}$ and *Stat3*$^{-/-}$ ESCs and suggests differences between network requirements for naive pluripotency establishment versus maintenance.

Active pSTAT3 lies downstream of a LIF-stimulated tyrosine kinase signaling cascade [1] (Figure S1A available online). Wild-type, *Nanog*$^{-/-}$, and constitutively NANOG-overexpressing ESCs were harvested from steady-state and LIF-induction cultures. Western blotting revealed that higher NANOG both increased steady-state pSTAT3 levels (Figures 1A and S1B) and enhanced ESC sensitivity to LIF stimulation (Figure 1B).

To explain how NANOG drives pSTAT3 elevation without directly binding the STAT3 protein or gene [17, 18], we examined the effect of NANOG on components of the LIF/STAT3 signaling pathway (Figure S1A). By quantitative RT-PCR (qRT-PCR), no correlation was reliably found between NANOG and transcript levels of the positive signal transducers *Lif*, *Lifr*, *Gp130*, *Jak2*, and *Stat3* (data not shown), despite NANOG binding to *Lif*, *Lifr*, and *Gp130* gene regulatory sequences in ESCs according to published chromatin immunoprecipitation sequencing data [18]. However, NANOG also binds the *Socs3* gene (Figure S1C), which is a negative regulator of STAT3 activation [19, 20]. This prompted our hypothesis that NANOG represses *Socs3* transcription.

Since *Socs3* transcription is upregulated by pSTAT3 to form a classic negative feedback loop [20, 21], determination of whether NANOG causes *Socs3* repression is obfuscated by the effect of pSTAT3 on *Socs3.* To disentangle the opposing yet interconnected influences of NANOG and pSTAT3 on *Socs3* expression, we designed experiments in which the pSTAT3 level is not significantly influenced by NANOG. ESCs can be maintained without exogenous LIF by 2i medium [10], which contains small molecules CHIR99021 (chiron) and PD0325901 (PD03) that inhibit GSK3 and MEK, respectively. In the absence of LIF-stimulated *Socs3* activation, the effect of NANOG on *Socs3* transcription can be assessed. From qRT-PCR analysis of wild-type, *Nanog*$^{-/-}$, and constitutively NANOG-overexpressing ESCs cultured in 2i, a strong negative correlation was evident between *Nanog* and *Socs3* expression levels (Figure 1C). Furthermore, in the absolute absence of pSTAT3 in *Stat3*$^{-/-}$ ESCs, those constitutively overexpressing NANOG exhibited a lower *Socs3* level (Figure 1C).

STAT3 can be specifically activated by GCSF stimulation of the GY118F receptor transgene [22–24] (Figure S1A). When GCSF/GY118F are used in the absence of LIF, *Socs3* downregulation by NANOG should have little or no effect on pSTAT3 levels; GY118F is insensitive to SOCS3 repression, and nearly all STAT3 activation will be attributable to GY118F rather than LIFR-GP130. In addition, we generated a doxycycline (dox)-inducible *Nanog* transgene (iNANOG) (Figure S1D). Dox induction of NANOG expression rapidly led to enriched NANOG binding at the *Socs3* gene in ESCs, consistent with direct transcriptional regulation (Figure S1C). In *Nanog*$^{-/-}$ ESCs containing both iNANOG and GY118F, NANOG expression and STAT3 activation were induced separately and in combination in 2i. Induction of NANOG alone caused *Socs3* repression, while *Socs3* induction in response to STAT3 activation was reduced by 60% when NANOG was also induced (Figure 1D). This suggests that *Socs3* repression is a mechanism by which NANOG

*Correspondence: jcs64@cscr.cam.ac.uk

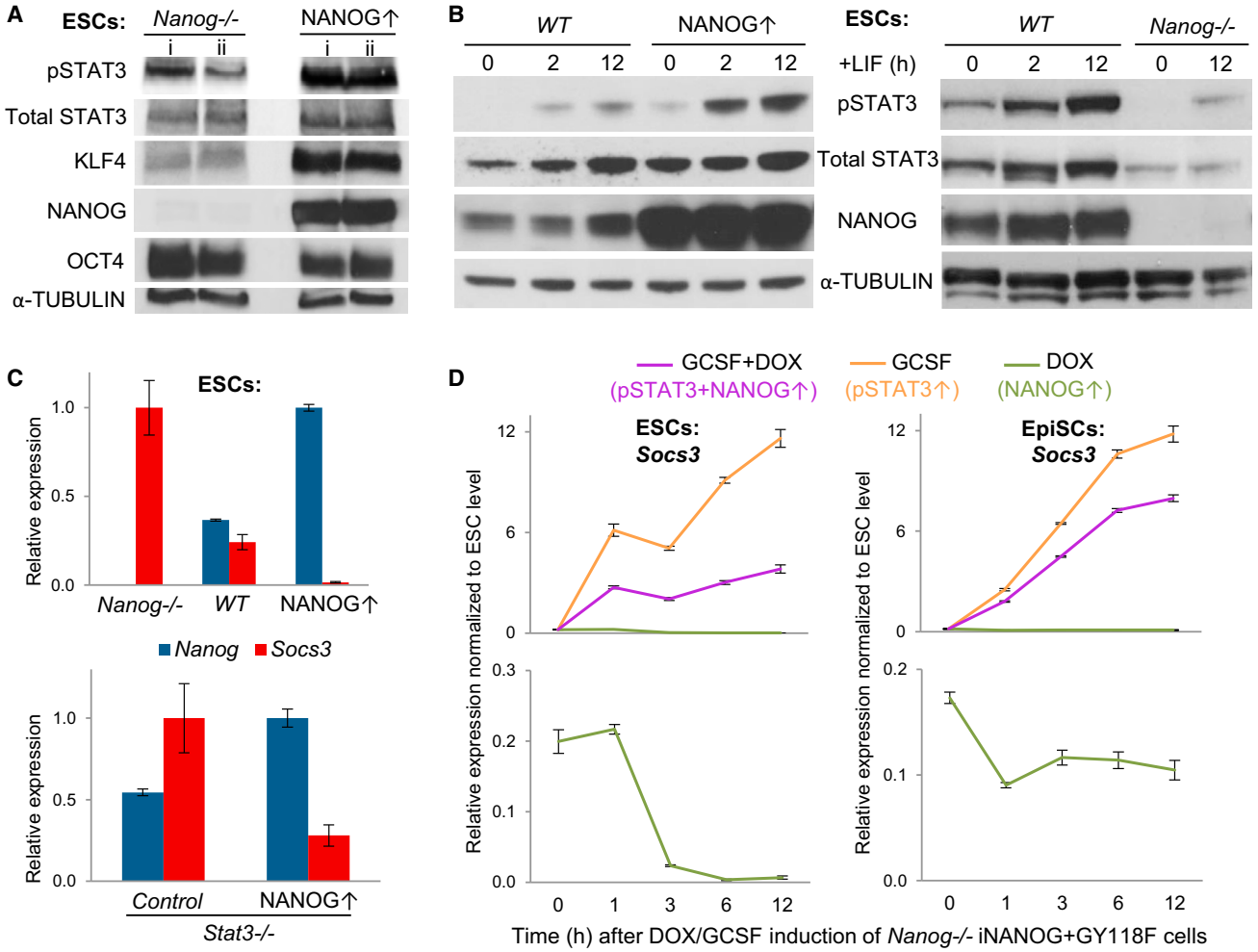


Figure 1. NANOG Amplifies STAT3 Activation

(A) Western blot analysis of pSTAT3, total STAT3, and KLF4 protein expression in *Nanog*$^{-/-}$ and constitutively NANOG-overexpressing ESCs cultured in steady-state serum+LIF conditions and harvested on two separate days (i and ii). OCT4 expression confirmed undifferentiated status.

(B) Western blot analysis of pSTAT3 and total STAT3 protein expression in wild-type, constitutively NANOG-overexpressing, and *Nanog*$^{-/-}$ ESCs in response to LIF stimulation for 0, 2, and 12 hr. Positive feedback of pSTAT3 induction on STAT3 and NANOG expression was observed, in accordance with previous work [2, 7, 15]. Selection for pluripotent cells was maintained throughout. Left: LIF was withdrawn from serum+LIF culture for 36 hr prior to readdition. Right: ESCs were cultured in 2i for at least 7 days prior to LIF addition. Different conditions were used to provide the most informative comparisons, since 2i boosts the expression of endogenous NANOG compared to serum conditions [16].

(C and D) qRT-PCR analysis of gene expression, relative to *Gapdh* and normalized either to the highest value (C) or to serum+LIF ESC level (D). Data shown are the mean of three technical replicates and are from one of two representative experiments. Error bars indicate ±SD.

(C) *Nanog* and *Socs3* expression in ESCs cultured in 2i without LIF for at least 7 days. Top: *Nanog*$^{-/-}$, wild-type, and constitutively NANOG-overexpressing ESCs. Bottom: control and NANOG-overexpressing *Stat3*$^{-/-}$ ESCs.

(D) Top left: *Socs3* expression in *Nanog*$^{-/-}$ iNANOG+GY118F ESCs after induction with dox and/or GCSF in 2i. Top right: *Socs3* expression in *Nanog*$^{-/-}$ iNANOG+GY118F EpiSCs after induction with dox and/or GCSF in FGF2+ActivinA. Bottom: zoomed-in view of the respective dox inductions.

See also Figure S1.

augments LIF signal transduction in ESCs, resulting in higher levels of active pSTAT3. The same trends were observed in *Nanog*$^{-/-}$ iNANOG+GY118F postimplantation epiblast stem cells (EpiSCs) in standard FGF2+ActivinA conditions (Figure 1D), demonstrating that NANOG-mediated *Socs3* repression is not restricted to ESCs and may be of functional relevance during NANOG-driven reprogramming.

NANOG and pSTAT3 Synergistically Upregulate KLF4

To further explore our newfound mechanistic link between NANOG and LIF/STAT3 signaling, we investigated the effect of NANOG on expression of LIF/STAT3 targets in ESCs. In the steady-state presence of LIF, we found strong positive correlation between levels of NANOG and KLF4, a canonical pSTAT3 target [2, 7, 8] (Figures 1A and S1B). Interestingly, in response to LIF stimulation, *Klf4* upregulation required NANOG to be present, while NANOG overexpression cooperated with LIF to substantially increase the rate and levels of *Klf4* induction (Figures 2A and 2B). In agreement with previous work [8], modest upregulation of NANOG was found in response to LIF (Figures 1B and S2A).

The correlation between NANOG and *Klf4* was abrogated in the absence of pSTAT3 in *Stat3*$^{+/+}$ ESCs without LIF and in *Stat3*$^{-/-}$ ESCs (Figure 2C), showing that NANOG-driven

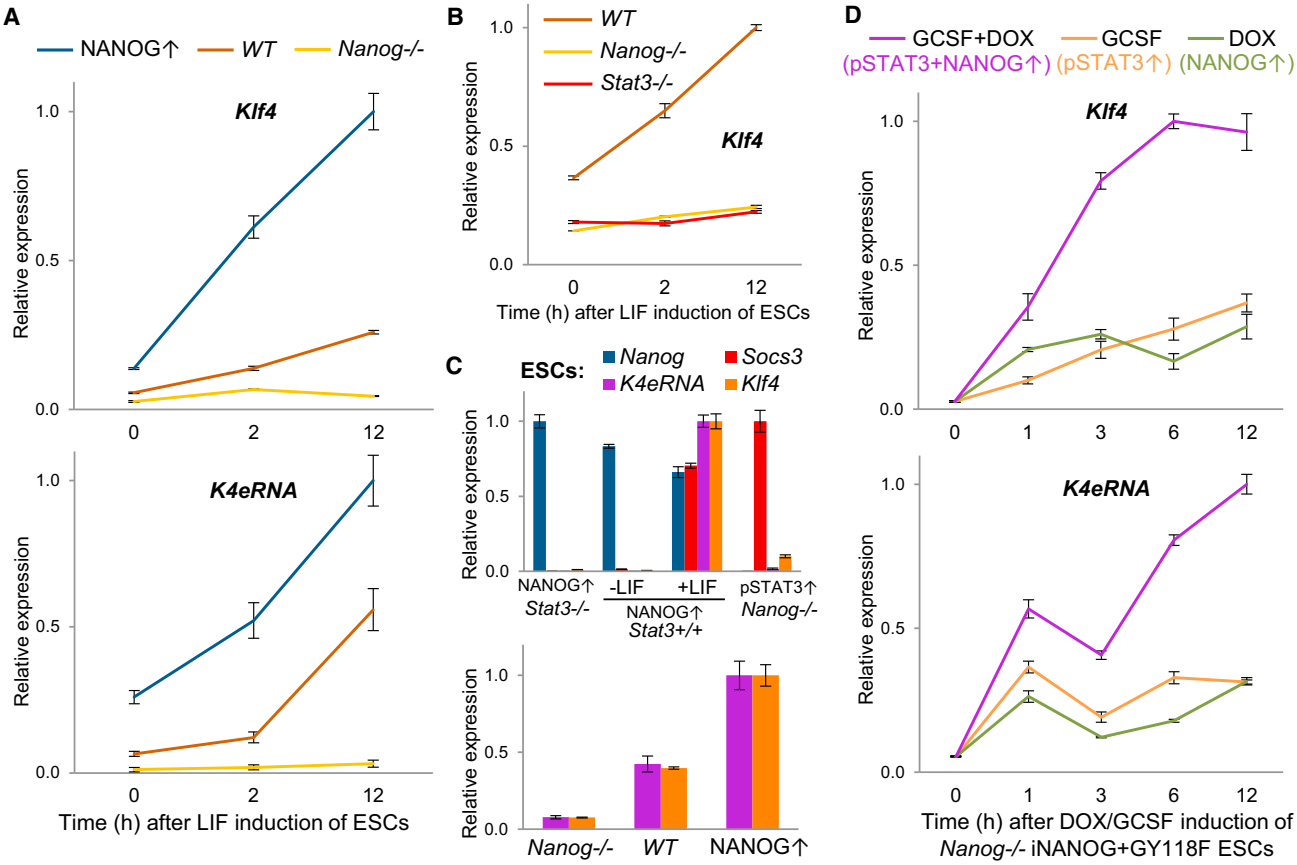


Figure 2. NANOG and pSTAT3 Synergistically Upregulate KLF4

qRT-PCR analysis of gene expression relative to *Gapdh* and normalized to the highest value. Data shown are the mean of three technical replicates and are from one of two representative experiments. Error bars indicate ±SD.

(A) *Klf4* and *K4eRNA* expression in *Nanog*$^{-/-}$, wild-type, and constitutively NANOG-overexpressing ESCs in response to LIF stimulation. LIF was withdrawn from serum culture for 36 hr prior to readdition. Selection for pluripotent cells was maintained throughout.

(B) *Klf4* expression in *Stat3*$^{-/-}$, *Nanog*$^{-/-}$, and wild-type ESCs in response to LIF stimulation. LIF was added after at least 7 days in 2i without LIF.

(C) Top: *Klf4* and *K4eRNA* expression in NANOG-overexpressing *Stat3*$^{-/-}$, NANOG-overexpressing *Stat3*$^{+/+}$, and STAT3-hyperactivated (GCSF/GY118F) *Nanog*$^{-/-}$ ESCs in basic conditions with or without LIF. *Socs3* expression indicates STAT3 activation as appropriate. Bottom: *Klf4* and *K4eRNA* expression in *Nanog*$^{-/-}$, wild-type, and constitutively NANOG-overexpressing ESCs cultured in steady-state serum+LIF.

(D) *Klf4* and *K4eRNA* expression in *Nanog*$^{-/-}$ iNANOG+GY118F ESCs after induction with dox and/or GCSF in 2i.

See also Figure S2.

Klf4 upregulation was pSTAT3 dependent. However, the effect of NANOG on *Klf4* transcription was not solely attributable to pSTAT3 elevation, since STAT3 hyperactivation in the absence of NANOG could not rescue *Klf4* expression (Figure 2C).

The relationship between NANOG, pSTAT3, and *Klf4* expression was further dissected using *Nanog*$^{-/-}$ iNANOG+GY118F ESCs in 2i without LIF. Together, induction of NANOG expression and STAT3 activation elicited *Klf4* upregulation in a synergistic manner compared to induction of either factor alone (Figure 2D). This synergistic action of NANOG and pSTAT3 is specific to *Klf4*: other pluripotency factors did not respond in this striking manner, including NANOG-target *Esrrb* [25] and pSTAT3-target *Klf5* [7, 8] (Figure S2B).

Since NANOG and pSTAT3 both bind the *Klf4* enhancer (Figure S2C), it is likely that they regulate *Klf4* transcription directly. The *Klf4* enhancer lies around 60 kb downstream of the *Klf4* transcription start site and was recently identified as an archetypal "super-enhancer" [26]. We found novel noncoding RNA to be expressed from the *Klf4* enhancer in ESCs, and we termed it *K4eRNA* (Figure S2C). Given that *K4eRNA* expression positively correlates with *Klf4*, responding to NANOG and LIF/STAT3 in the same synergistic manner (Figures 2A, 2C, and 2D), we hypothesize that *K4eRNA* is a *cis* activator of *Klf4* transcription.

NANOG and pSTAT3 Induce Rapid and Efficient Reactivation of Naive Genes

EpiSC reprogramming requires reversion from primed to naive pluripotency and thus provides an excellent system in which to study naive pluripotency acquisition. Conversion of EpiSCs to iPSCs does not occur simply in naive-state culture conditions, but can be driven by a minimum of one factor [27]. It is known that NANOG overexpression and STAT3 hyperactivation together increase EpiSC reprogramming efficiency in a synergistic manner [1]. Since we have mechanistically linked NANOG and pSTAT3 in ESCs where the pluripotency network is fully operational, we turned to EpiSCs to study the role of these connected mechanisms during naive pluripotency establishment.

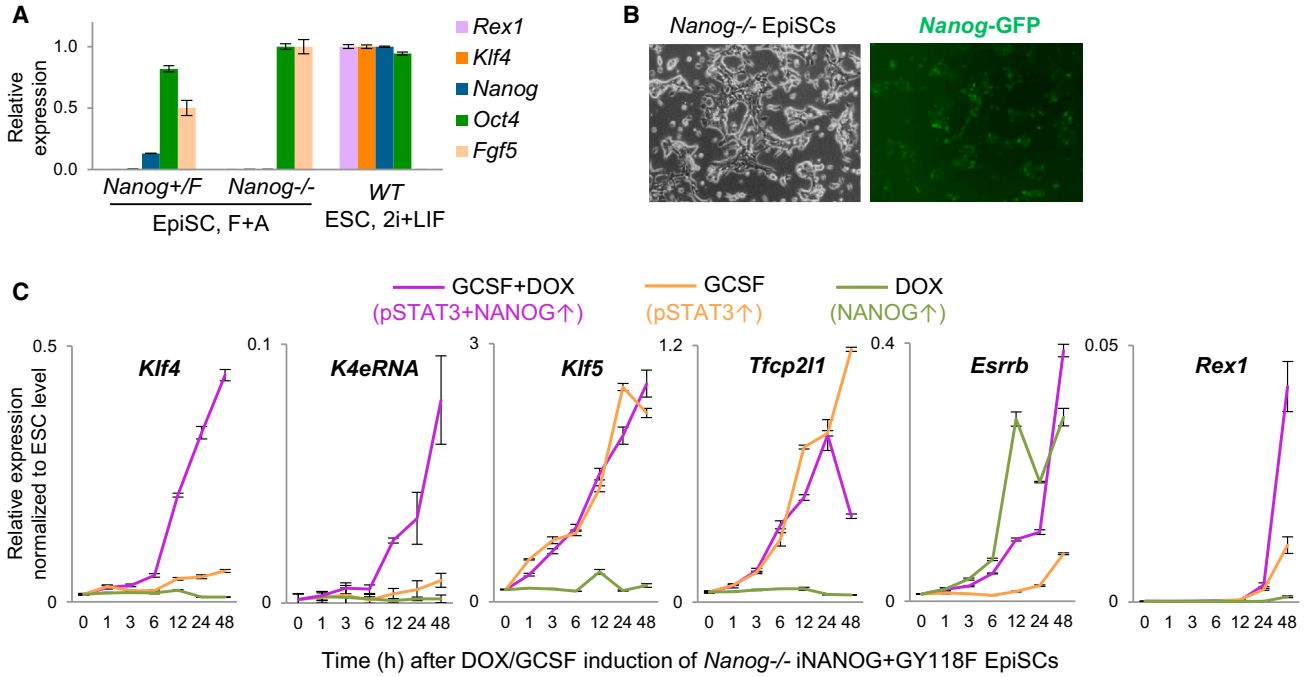


Figure 3. NANOG and pSTAT3 Induce Rapid and Efficient Reactivation of Naive Genes

(A) qRT-PCR analysis of *Nanog*$^{-/-}$ and *Nanog*$^{+/Flox}$ (*Nanog*$^{+/F}$) EpiSCs derived from *Nanog*$^{-/F}$ and *Nanog*$^{+/F}$ littermate embryos, cultured in FGF2+ActivinA (F+A), compared to wild-type ESCs in 2i+LIF. Lack of *Nanog* expression in *Nanog*$^{-/-}$ EpiSCs confirmed the null genotype, while their expression of *Oct4* and *Fgf5* but not *Rex1* or *Klf4* confirmed their EpiSC identity. Gene expression was measured relative to *Gapdh* and normalized to the highest value. Data shown are the mean of three technical replicates and are from one of two representative experiments. Error bars indicate ±SD.

(B) Representative phase and *Nanog*-GFP images of the *Nanog*$^{-/-}$ EpiSCs derived from *Nanog*$^{-/F}$ embryos, in FGF2+ActivinA conditions. GFP reporter is under the endogenous *Nanog* promoter of the floxed null allele (Figure S3A). Images are 1122 μm by 839 μm.

(C) qRT-PCR analysis of *Klf4*, *K4eRNA*, *Klf5*, *Tfcp2l1*, *Esrrb*, and *Rex1* expression in *Nanog*$^{-/-}$ iNANOG+GY118F EpiSCs after induction with dox and/or GCSF in FGF2+ActivinA. Gene expression was measured relative to *Gapdh* and normalized to serum+LIF ESC level = 1. Data shown are the mean of three technical replicates and are from one of two representative experiments. Error bars indicate ±SD.

See also Figure S3.

The iNANOG+GY118F system provides a powerful platform for the quantitative dissection of NANOG and pSTAT3 mechanisms, as they can be induced separately and in combination within a single cell line. We generated *Nanog*$^{-/-}$ background EpiSCs (Figure 3A, 3B, and S3A–S3C) to eliminate confounding endogenous *Nanog* expression, and maintained EpiSC FGF2+ActivinA culture conditions so that putative reprogramming kinetics could be ascribed exclusively to transgene induction. Strikingly, NANOG and pSTAT3 codependently reactivated *Klf4* and *K4eRNA* in this distinct cellular and environmental context (Figure 3C). This demonstrates that their effect on *Klf4* is not an ESC-specific phenomenon and may be of functional relevance for NANOG/pSTAT3-driven reprogramming. We observed activation of naive pluripotency marker *Rex1* at 48 hr (Figure 3C). Since *Rex1* did not respond to NANOG/pSTAT3 in this manner in ESCs (Figure S2B), we believe *Rex1* induction in EpiSCs to be indirect, indicating identity changes toward iPSCs within the population. Interestingly, *Rex1* activation positively correlates with *Klf4* expression.

Known NANOG-target *Esrrb* [25] responded to dox induction of NANOG expression in EpiSCs (Figure 3C). Similarly, known pSTAT3 targets *Klf5* [7, 8] and *Tfcp2l1* [15, 28] were upregulated after GCSF induction of STAT3 activation (Figure 3C). However, the synergistic response of *Klf4* to NANOG and pSTAT3 remains unique. In total, NANOG and pSTAT3 rapidly reactivated many key components of the naive pluripotency network to near ESC level, shedding light on their ability to drive fast and efficient reprogramming. This is even more remarkable when taking into account that the assay used EpiSC culture conditions instead of conditions promoting reprogramming or ESC self-renewal.

Combined pSTAT3 and KLF4 Bypass NANOG in Reprogramming

Although NANOG is dispensable for pluripotency maintenance [11], it is required for establishment of the pluripotent epiblast during preimplantation embryonic development [4]. Correspondingly, NANOG is essential for naive pluripotency establishment during conventional in vitro reprogramming experiments [4]. Rescue of *Nanog*$^{-/-}$ reprogramming thus provides a means of functionally testing proposed downstream mechanisms of NANOG.

We have described two new NANOG mechanisms: pSTAT3 elevation by SOCS3 repression and KLF4 upregulation in cooperation with pSTAT3. Therefore, we investigated the ability of pSTAT3 and KLF4 to rescue reprogramming of *Nanog*$^{-/-}$ EpiSCs. We also tested ESRRB, since it has previously been reported as a NANOG downstream target able to bypass NANOG in reprogramming of ESC-derived EpiSCs [25]. Expression of *GFP* and *βgeo* under the control of endogenous *Nanog* promoters provide visual and selective reporters in our system, since EpiSCs cannot survive long-term in the reprogramming culture conditions (Figure 4A).

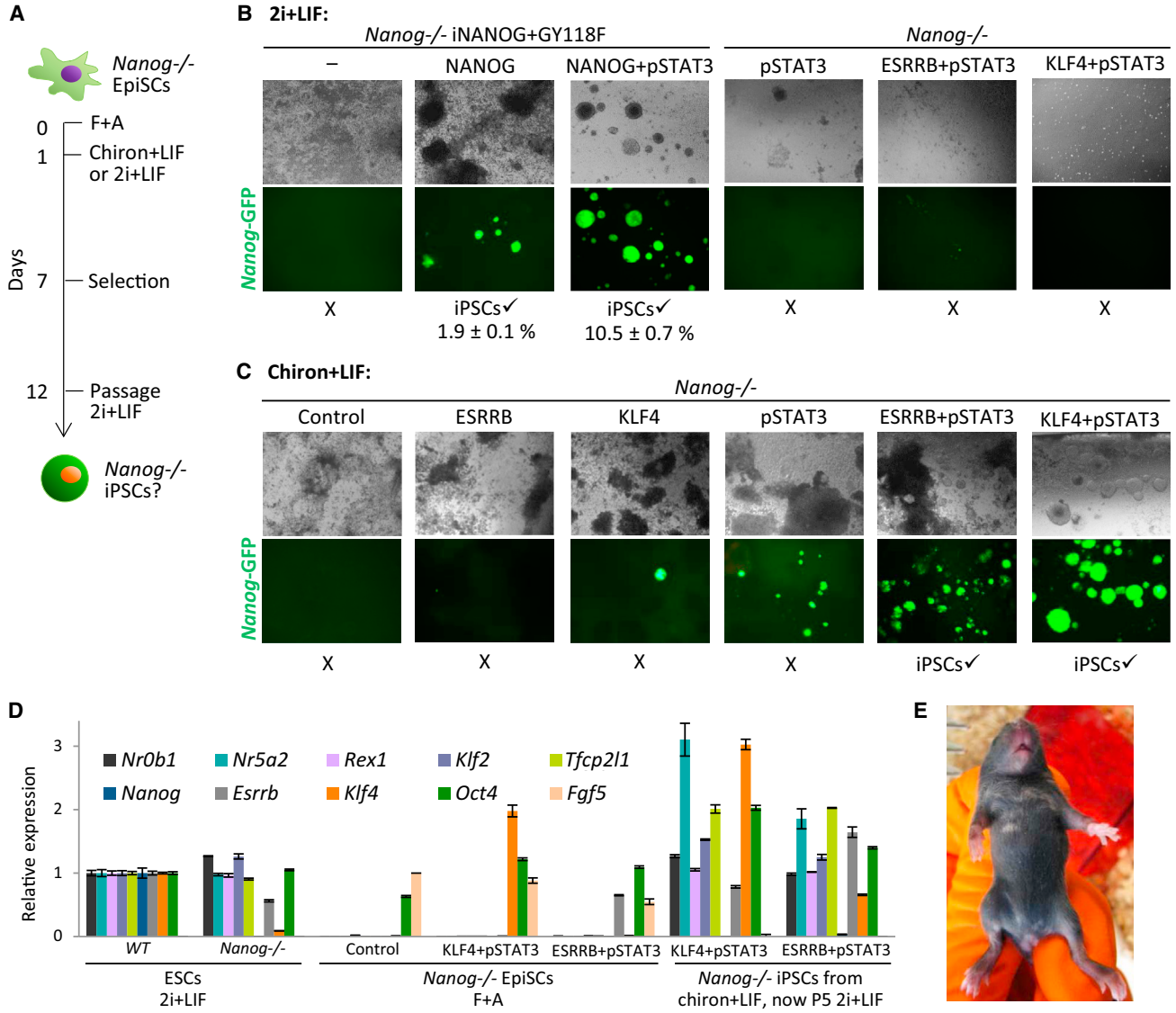


Figure 4. Combined pSTAT3 and KLF4 Bypass NANOG in Reprogramming

(A) Schematic depicting the protocol for reprogramming of *Nanog*$^{-/-}$ EpiSCs containing *βgeo* and *GFP* reporters under the endogenous *Nanog* promoters (Figure S3A). EpiSCs were plated in FGF2+ActivinA (F+A) and, after 1 day, medium was switched to chiron+LIF or 2i+LIF to prompt reprogramming. On day 7, G418, FGFR inhibitor, and ALK inhibitor were added to select for emergent iPSCs. On day 12, iPSCs were passaged into 2i+LIF.

(B) Representative phase and *Nanog*-GFP images taken on day 12 of reprogramming in 2i+LIF. The mean number of GFP$^+$ iPSC colonies is indicated as the percentage of cells initially plated, ±SD (n = 3 biological replicates). *Nanog*$^{-/-}$ iNANOG+GY118F EpiSCs generated GFP$^+$ iPSCs in 2i+LIF only if dox or dox+GCSF were supplied during reprogramming, to induce NANOG expression and STAT3 activation, respectively. *Nanog*$^{-/-}$ EpiSCs with activated STAT3 (GCSF/GY118F) and constitutive ESRRB or KLF4 expression did not generate GFP$^+$ iPSCs in 2i+LIF.

(C) Representative phase and *Nanog*-GFP images taken on day 12 of reprogramming in chiron+LIF. Quantification of reprogramming efficiency by GFP$^+$ colony counting was inappropriate in this case, since chiron+LIF is a permissive medium compared to 2i+LIF and yielded some GFP$^+$ colonies that were not true iPSCs. After passaging in 2i+LIF, clean GFP$^+$ iPSC lines were obtained as indicated.

(D) qRT-PCR analysis of gene expression in chiron+LIF-derived *Nanog*$^{-/-}$ iPSCs after passaging in 2i+LIF, compared to parental EpiSC lines in FGF2+ActivinA (F+A) and wild-type and *Nanog*$^{-/-}$ ESCs in 2i+LIF. Expression patterns of *Nanog*, *Esrrb*, and *Klf4* verified the genotypes. Reprogramming to naive iPSCs was confirmed by reactivation of naive genes, maintenance of *Oct4*, and repression of *Fgf5* expression. Gene expression was measured relative to *Gapdh* and normalized to wild-type ESC level, except *Fgf5*, which was normalized to control EpiSCs. Data shown are the mean of three technical replicates and are from one of two representative experiments. Error bars indicate ±SD.

(E) Chimeric mouse obtained after injection of pSTAT3+KLF4 EpiSC-derived *Nanog*$^{-/-}$ iPSCs into C57/BL6 blastocysts. Agouti coat color indicates chimeric contribution (MF1/129 background).

See also Figure S4.

Forced expression of NANOG, ESRRB, KLF4 or pSTAT3 individually can drive reprogramming of *Nanog*$^{+/+}$ EpiSCs in 2i+LIF conditions [1, 4, 25, 27]. We verified the ability of our *Nanog*$^{-/-}$ EpiSCs to generate iPSCs in 2i+LIF when rescued by NANOG expression and confirmed that STAT3 activation in conjunction with NANOG expression drives rapid reprogramming at high efficiency (Figure 4B). However, in the absence of NANOG, individual and combined overexpression

of ESRRB, KLF4, and pSTAT3 could not rescue reprogramming in 2i+LIF (Figure 4B).

In parallel, we tried to rescue *Nanog*$^{-/-}$ reprogramming with individual and combined overexpression of ESRRB, KLF4, and pSTAT3 in chiron+LIF (Figure 4C) and in serum+LIF (data not shown). To our surprise, we found that ESRRB overexpression was unable to drive *Nanog*$^{-/-}$ EpiSC reprogramming in any condition. We speculate that this is due to differences in reprogramming propensity between our EpiSCs derived from *Nanog*$^{-/F}$ embryos and ESC-derived secondary EpiSC systems [25]. Individually, KLF4 and pSTAT3 also failed to yield iPSCs. Although they initially generated GFP^{+} colonies in chiron+LIF (Figure 4C), these lacked the capacity to self-renew after passaging into 2i+LIF, suggesting that reprogramming was incomplete.

In combination, pSTAT3+ESRRB and pSTAT3+KLF4 yielded *Nanog*$^{-/-}$ GFP^{+} iPSCs in chiron+LIF conditions (Figure 4C). Reprogramming occurred with the highest rate and efficiency with pSTAT3+KLF4. Although *Nanog*$^{-/-}$ iPSCs could not be established in 2i+LIF, once naive pluripotency was established in chiron+LIF, *Nanog*$^{-/-}$ iPSCs could be passaged indefinitely in 2i+LIF, consistent with known discrepancies between NANOG requirement in pluripotency establishment versus maintenance. After passaging in 2i+LIF, the gene expression profile and observed chimeric competence of *Nanog*$^{-/-}$ iPSCs formally demonstrated their reacquisition of a naive pluripotent program (Figures 4D and 4E). This highlights pSTAT3 activation as a key functional mechanism acting downstream of NANOG, which, in conjunction with overexpression of either NANOG-target KLF4 or ESRRB, can efficiently rescue *Nanog*$^{-/-}$ reprogramming. It is of interest to note that KLF4 but not NANOG overexpression can drive EpiSC reprogramming in the presence of JAK inhibitor [1], further supporting the placement of NANOG upstream and KLF4 downstream of STAT3 activation during reprogramming.

NANOG is dispensable for the initial formation of reprogramming intermediates (pre-iPSCs) from somatic cells, but is essential for pre-iPSCs to transit to naive pluripotency in 2i+LIF [4] (confirmed in Figures S4A and S4B). Since *Nanog*$^{-/-}$ EpiSCs were able to reprogram in chiron+LIF but not 2i+LIF, we tested whether *Nanog*$^{-/-}$ pre-iPSCs could also reprogram in alternative conditions. From these, we successfully obtained *Nanog*$^{-/-}$ iPSCs in chiron+LIF and KSR+LIF conditions, albeit with low speed and efficiency (Figures S4C–S4E). Microarray analysis revealed that pre-iPSC-derived and EpiSC-derived *Nanog*$^{-/-}$ iPSCs clustered closely with both wild-type and *Nanog*$^{-/-}$ ESCs, demonstrating reprogramming to a naive pluripotent identity (Figure S4F). It should be noted that retroviral *Klf4* and exogenous LIF provided the reprogramming impetus for these pre-iPSCs, again implicating pSTAT3 and KLF4 in the rescue of *Nanog*$^{-/-}$ reprogramming.

Discussion

We connect NANOG with the activation of STAT3, two major mechanisms for the establishment and maintenance of naive pluripotency. Our finding that NANOG modulates signal transduction of extracellular cues adds a new dimension to the interplay between external environment and nuclear control networks to instate and reinforce cellular identity. We also provide mechanistic insight into the process of induced pluripotency by showing that expression of KLF4, a canonical "Yamanaka" factor, results from codependent and synergistic action between NANOG and pSTAT3. Interestingly, the only remaining factor to be used by all reprogramming protocols is LIF and consequently STAT3 activation, now linked to NANOG.

The role of NANOG is thus to build a naive pluripotent transcriptional network by concurrently inducing the expression of ESRRB, enhancing LIF/STAT3 signal transduction, and inducing KLF4 expression in cooperation with pSTAT3 (Figure S4G). Ultimately, combinations of these factors allow in vitro bypassing of NANOG for the establishment of a naive pluripotent cell state. However, the observed slower kinetics and reduced efficiency of *Nanog*$^{-/-}$ somatic cell reprogramming imply the existence of further mechanisms by which NANOG operates. These may include additional downstream effectors of reprogramming and the activities of NANOG cofactors such as TET1/2 [29]. In this light, it is interesting to note that bypass of NANOG in reprogramming was enhanced by KSR medium (Figure S4C), which contains ascorbic acid, a powerful coactivator of dioxygenases such as the jumonji histone demethylases and TETs [30–34]. Additionally, it will be of future interest to ascertain why 2i conditions are detrimental to *Nanog*$^{-/-}$ reprogramming.

Successful induction of naive pluripotency can be achieved by the combined actions of different culture environments with different sets of transgenes. This has led to the notion that iPSCs can be generated by different, independent reprogramming mechanisms acting in an additive, linear manner. In contrast to this, our newfound cooperative relationship between NANOG and STAT3 activity raises the possibility of an integrated reprogramming mechanism. Therefore, we propose that allegedly independent mechanisms of naive pluripotency induction may instead be linked and centered on a small group of genes including NANOG.

Accession Numbers

The Gene Expression Omnibus accession number for the microarray data reported in this paper is GSE53529.

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2013.12.040.

Acknowledgments

We thank William Mansfield and Charles-Etienne Dumeau for blastocyst injections and Jianlong Wang for the inducible *Nanog* transgene. We are grateful to Rodrigo Santos for technical assistance, Annie Hoxhalli for genotyping advice, and Yael Costa and Austin Smith for discussions and reagents. This study was supported by Wellcome Trust Fellowships (WT086692 and WT101861). J.C.R.S. is a Wellcome Trust Research Fellow. H.T.S. is a recipient of an MRC Ph.D. studentship.

Received: October 18, 2013
Revised: December 5, 2013
Accepted: December 13, 2013
Published: January 23, 2014

References

1. Yang, J., van Oosten, A.L., Theunissen, T.W., Guo, G., Silva, J.C., and Smith, A. (2010). Stat3 activation is limiting for reprogramming to ground state pluripotency. Cell Stem Cell *7*, 319–328.
2. van Oosten, A.L., Costa, Y., Smith, A., and Silva, J.C. (2012). JAK/STAT3 signalling is sufficient and dominant over antagonistic cues for the establishment of naive pluripotency. Nat Commun *3*, 817.

3. Theunissen, T.W., van Oosten, A.L., Castelo-Branco, G., Hall, J., Smith, A., and Silva, J.C. (2011). Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. Curr. Biol. *21*, 65–71.
4. Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737.
5. Silva, J., Chambers, I., Pollard, S., and Smith, A. (2006). Nanog promotes transfer of pluripotency after cell fusion. Nature *441*, 997–1001.
6. Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.
7. Bourillot, P.Y., Aksoy, I., Schreiber, V., Wianny, F., Schulz, H., Hummel, O., Hubner, N., and Savatier, P. (2009). Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. Stem Cells *27*, 1760–1771.
8. Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118–122.
9. Do, D.V., Ueda, J., Messerschmidt, D.M., Lorthongpanich, C., Zhou, Y., Feng, B., Guo, G., Lin, P.J., Hossain, M.Z., Zhang, W., et al. (2013). A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. Genes Dev. *27*, 1378–1390.
10. Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.
11. Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.
12. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642.
13. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643–655.
14. Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., and Yokota, T. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. EMBO J. *18*, 4261–4269.
15. Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J. *32*, 2561–2574.
16. Hamazaki, T., Kehoe, S.M., Nakano, T., and Terada, N. (2006). The Grb2/Mek pathway represses Nanog in murine embryonic stem cells. Mol. Cell. Biol. *26*, 7539–7549.
17. Torres, J., and Watt, F.M. (2008). Nanog maintains pluripotency of mouse embryonic stem cells by inhibiting NFkappaB and cooperating with Stat3. Nat. Cell Biol. *10*, 194–201.
18. Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.
19. Kershaw, N.J., Murphy, J.M., Liau, N.P., Varghese, L.N., Laktyushin, A., Whitlock, E.L., Lucet, I.S., Nicola, N.A., and Babon, J.J. (2013). SOCS3 binds specific receptor-JAK complexes to control cytokine signaling by direct kinase inhibition. Nat. Struct. Mol. Biol. *20*, 469–476.
20. Schmitz, J., Weissenbach, M., Haan, S., Heinrich, P.C., and Schaper, F. (2000). SOCS3 exerts its inhibitory function on interleukin-6 signal transduction through the SHP2 recruitment site of gp130. J. Biol. Chem. *275*, 12848–12856.
21. Yoshiura, S., Ohtsuka, T., Takenaka, Y., Nagahara, H., Yoshikawa, K., and Kageyama, R. (2007). Ultradian oscillations of Stat, Smad, and Hes1 expression in response to serum. Proc. Natl. Acad. Sci. USA *104*, 11292–11297.
22. Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev. *12*, 2048–2060.
23. Burdon, T., Chambers, I., Stracey, C., Niwa, H., and Smith, A. (1999). Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. Cells Tissues Organs (Print) *165*, 131–143.
24. Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Dev. Biol. *210*, 30–43.
25. Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell *11*, 477–490.
26. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.
27. Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development *136*, 1063–1069.
28. Ye, S., Li, P., Tong, C., and Ying, Q.L. (2013). Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. EMBO J. *32*, 2548–2560.
29. Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370–374.
30. Esteban, M.A., and Pei, D. (2012). Vitamin C improves the quality of somatic cell reprogramming. Nat. Genet. *44*, 366–367.
31. Monfort, A., and Wutz, A. (2013). Breathing-in epigenetic change with vitamin C. EMBO Rep. *14*, 337–346.
32. Yin, R., Mao, S.Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., et al. (2013). Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. J. Am. Chem. Soc. *135*, 10396–10403.
33. Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature *500*, 222–226.
34. Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y., et al. (2013). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat. Genet. *45*, 34–42.

Current Biology, Volume *24*

Supplemental Information

NANOG Amplifies STAT3 Activation and They Synergistically Induce the Naive Pluripotent Program

Hannah T. Stuart, Anouk L. van Oosten, Aliaksandra Radzisheuskaya, Graziano Martello, Anzy Miller, Sabine Dietmann, Jennifer Nichols, and José C.R. Silva

Figure S1:

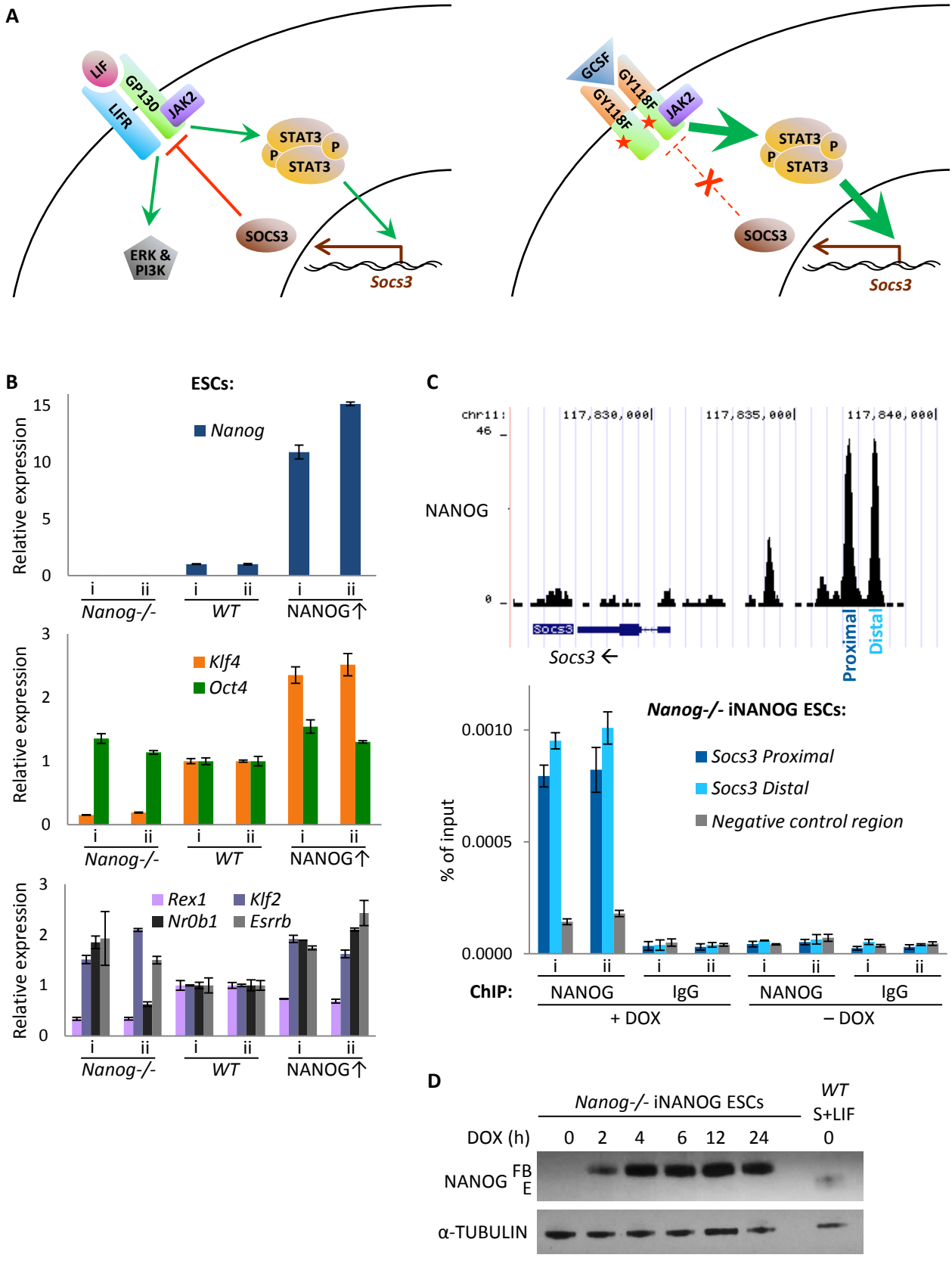

A
LIF
GP130
LIFR
JAK2
STAT3
P
STAT3
SOCS3
ERK &
PI3K
Socs3
GCSF
GY118F
GY118F
JAK2
STAT3
P
STAT3
SOCS3
Socs3
B
ESCs:
Nanog
Relative expression
15
10
5
0
i
ii
Nanog-/-
WT
NANOG↑
Klf4
Oct4
3
2
1
0
Rex1
Klf2
Nr0b1
Esrrb
C
chr11:
46
117,830,000
117,835,000
117,840,000
NANOG
Socs3
Socs3 ←
Proximal
Distal
Nanog-/- iNANOG ESCs:
Socs3 Proximal
Socs3 Distal
Negative control region
% of input
0.0010
0.0005
0.0000
ChIP:
NANOG
IgG
+ DOX
– DOX
D
Nanog-/- iNANOG ESCs
WT
S+LIF
DOX (h)
0
2
4
6
12
24
0
NANOG
FBE
α-TUBULIN

Figure S1: NANOG amplifies STAT3 activation (Related to Figure 1)
(A) Simplified schematic comparing LIF/LIFR-GP130 and GCSF/GY118F signal transduction pathways. Left: LIF ligand binds the extracellular domains of LIFR and GP130, triggering phosphorylation and activation of JAK2 associated with the intracellular domain of GP130. JAK2 phosphorylates the intracellular domain of GP130, causing recruitment of latent cytoplasmic STAT3 to GP130 via the STAT3 SH2 phosphotyrosine-binding domain. JAK2 then phosphorylates tyrosine-705 of STAT3 (pSTAT3), allowing SH2-mediated homodimerization of pSTAT3, nuclear translocation and transcriptional activation of multiple target genes [S1-5]. *Socs3* transcription is rapidly and strongly induced by pSTAT3. SOCS3 binds to GP130-JAK2 complexes, prevents STAT3 binding to JAK2, and thus inhibits STAT3 phosphorylation and activation [S6, S7], forming a classic negative feedback loop in which pSTAT3 and SOCS3 levels oscillate in antiphase [S8]. Additionally, LIF signalling leads to activation of the MAPK/ERK and PI3-kinase pathways [S9, S10]. Right: Specific and sustained STAT3 activation can be achieved by supplying GCSF to cells expressing GY118F, a chimeric transmembrane receptor [S11-13]. The GY118F extracellular domain is that of the human GCSF receptor, whereas the transmembrane and intracellular domains are based on murine GP130, with a point mutation of tyrosine-118 to phenylalanine. The Y118F mutation prevents activation of MAPK/ERK and PI3-kinase pathways, and blocks SOCS3-mediated negative feedback by eliminating the SOCS3 binding site on GP130.
(B) RT-qPCR analysis of gene expression in *Nanog-/-*, wild-type and constitutively NANOG-overexpressing ESCs cultured in steady-state serum+LIF conditions and harvested on 2 separate days (i and ii). Expression was measured relative to *Gapdh* and normalized to the appropriate wild-type sample. *Oct4* expression confirmed undifferentiated status. *Rex1, Klf2,* and *Nr0b1* expression patterns confirmed that the strong positive correlation observed between *Nanog* and *Klf4* was specific. Data shown are the mean of 3 technical replicates. Error bars indicate ± s.d.
(C) Upper: Published ESC ChIP-seq data [S14] reveal that NANOG binds the upstream regulatory region of the *Socs3* gene. Arrow indicates the direction of *Socs3* transcription. Lower: ChIP analysis of NANOG binding at the regulatory region of *Socs3*. Proximal and distal NANOG binding sites are as indicated above. NANOG expression was dox-induced in *Nanog-/-* iNANOG ESCs in 2i culture and samples were harvested at t=0 and 3 hours. Data shown are the mean of 3 technical replicates from each of 2 experiments (i and ii). Error bars indicate ± s.d.
(D) Western blot analysis of NANOG protein expression in *Nanog-/-* iNANOG ESCs following dox induction in serum+LIF, compared to wild-type ESCs cultured in serum+LIF. FLAG- and biotin-tagged (FB) NANOG expressed from iNANOG was detected at a higher molecular weight than endogenous (E) NANOG.

Figure S2:

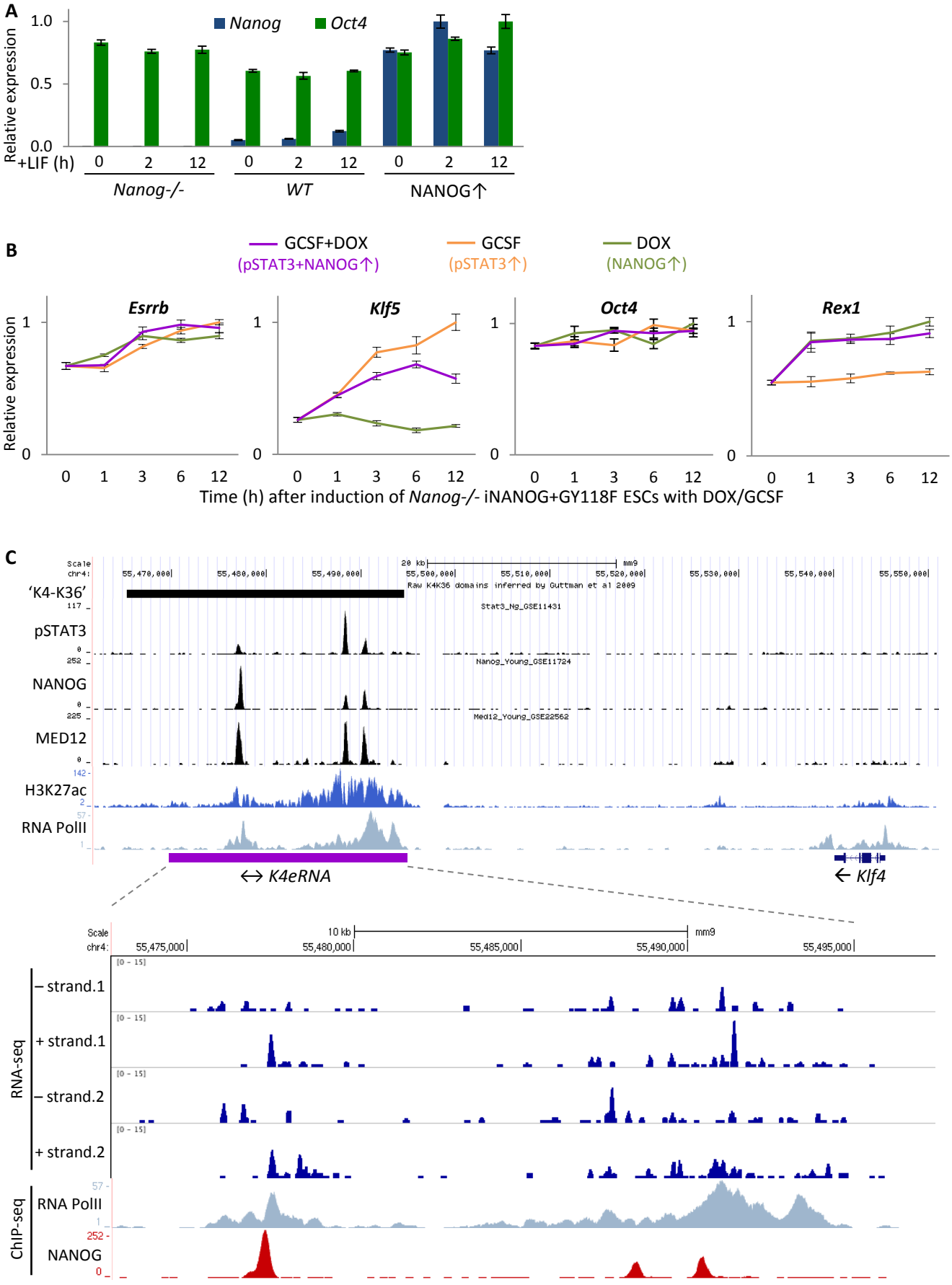

A
Relative expression
1.0
0.5
0.0
Nanog
Oct4
+LIF (h)
0
2
12
0
2
12
0
2
12
Nanog-/-
WT
NANOG↑
B
GCSF+DOX
(pSTAT3+NANOG↑)
GCSF
(pSTAT3↑)
DOX
(NANOG↑)
Esrrb
Klf5
Oct4
Rex1
Relative expression
1
0
0
1
3
6
12
Time (h) after induction of Nanog-/- iNANOG+GY118F ESCs with DOX/GCSF
C
Scale
chr4:
55,470,000
55,480,000
55,490,000
20 kb
55,500,000
55,510,000
55,520,000
mm9
55,530,000
55,540,000
55,550,000
'K4-K36'
Raw K4K36 domains inferred by Guttman et al 2009
Stat3_Ng_GSE11431
pSTAT3
Nanog_Young_GSE11724
NANOG
Med12_Young_GSE22562
MED12
H3K27ac
RNA PolII
↔ K4eRNA
← Klf4
Scale
chr4:
55,475,000
10 kb
55,480,000
55,485,000
55,490,000
mm9
55,495,000
RNA-seq
– strand.1
+ strand.1
– strand.2
+ strand.2
ChIP-seq
RNA PolII
NANOG

Figure S2: NANOG and pSTAT3 synergistically upregulate KLF4 (Related to Figure 2)
(A) RT-qPCR analysis of *Oct4* and *Nanog* expression, relative to *Gapdh* and normalized to the highest value. LIF was withdrawn from serum culture of *Nanog-/-*, wild-type and constitutively NANOG-overexpressing ESCs for 36 hours prior to re-addition. Selection for pluripotent cells was maintained throughout, and *Oct4* expression indicated undifferentiated status. Data shown are the mean of 3 technical replicates and are from 1 of 2 representative experiments. Error bars indicate ± s.d.
(B) RT-qPCR analysis of *Esrrb, Klf5, Oct4* and *Rex1* expression in *Nanog-/-* iNANOG+GY118F ESCs following induction with dox and/or GCSF in 2i. Gene expression was measured relative to *Gapdh* and normalized to the highest value. Data shown are the mean of 3 technical replicates and are from 1 of 2 representative experiments. Error bars indicate ± s.d.
(C) The *Klf4* gene locus shown on assembly mm9. Upper: Analysis of published ESC ChIP-seq data revealed pSTAT3, NANOG and Mediator (MED12) binding to the *Klf4* downstream enhancer. RNA PolII binding and H3K27ac enrichment suggest active transcription from this enhancer. Furthermore, a 'K4-K36 domain' is predicted here, which, in the absence of a known protein-coding gene, is thought to demarcate PolII-transcribed non-coding RNA [S15]. We found novel non-coding RNA to be expressed from the *Klf4* enhancer in ESCs, termed *K4eRNA,* the position of which is marked by the purple box. Arrows indicate the directions of *K4eRNA* and *Klf4* transcription. Lower: Published RNA-seq [S16] indicates bidirectional transcription from the *Klf4* downstream enhancer in wild-type ESCs in 2i+LIF, confirmed here by RT-PCR (data not shown).

Figure S3:

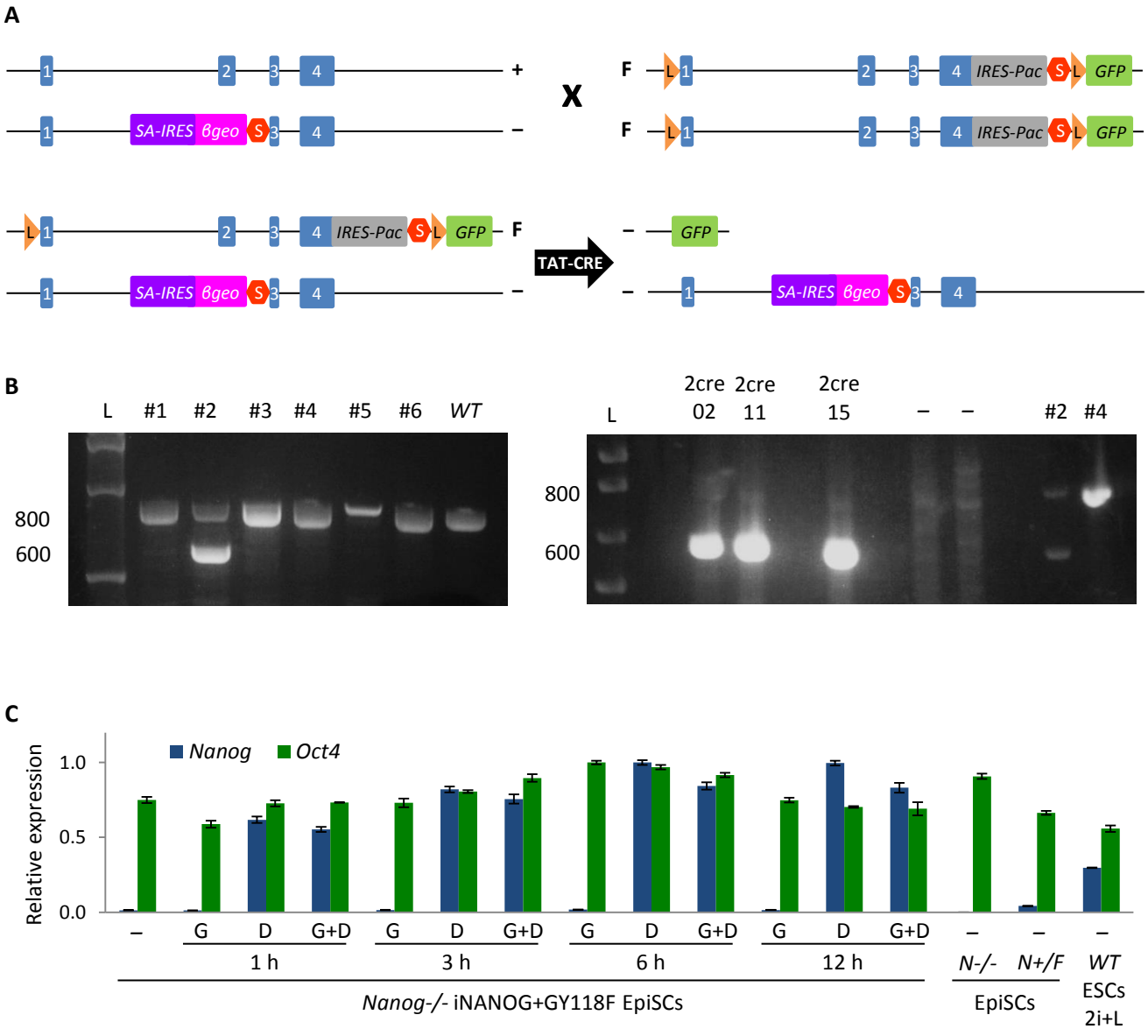

A
SA-IRES
βgeo
IRES-Pac
GFP
TAT-CRE
B
L
#1
#2
#3
#4
#5
#6
WT
800
600
2cre 02
2cre 11
2cre 15
C
Nanog
Oct4
Relative expression
1.0
0.5
0.0
G
D
G+D
1 h
3 h
6 h
12 h
N-/-
N+/F
WT ESCs 2i+L
Nanog-/- iNANOG+GY118F EpiSCs
EpiSCs

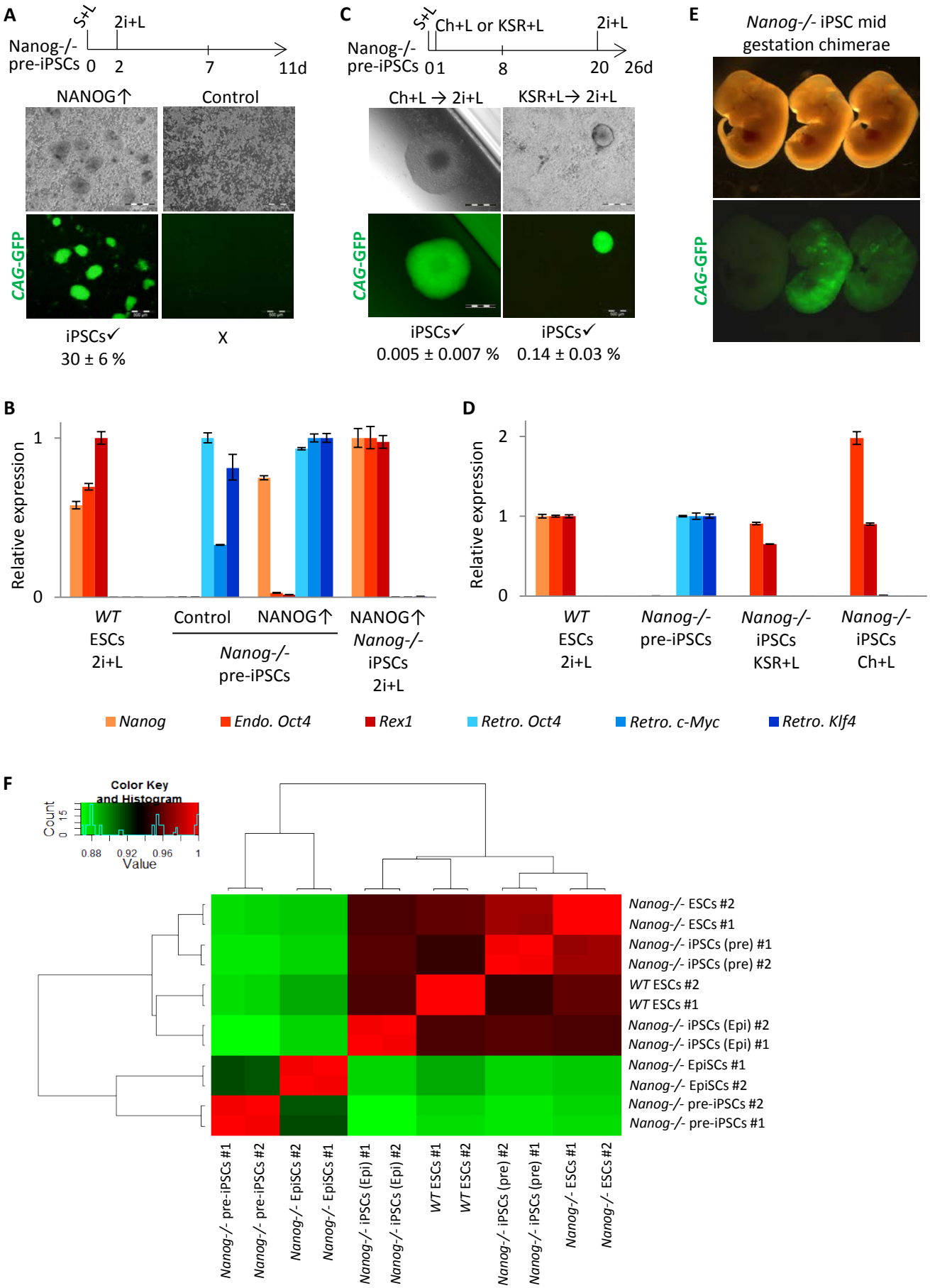
Figure S3: NANOG and pSTAT3 induce rapid and efficient reactivation of naïve genes (Related to Figure 3)

(A) Simplified schematic of the strategy used to generate *Nanog-/-* EpiSCs, not to scale. All alleles are under control of the endogenous *Nanog* promoter, and the *Nanog* 5' UTR is unchanged. We crossed *NanogF/F* [S17] and *Nanog+/-* [S18] mice and derived EpiSCs from the resulting *Nanog-/F* and *Nanog+/F* post-implantation embryos. We treated *Nanog-/F* EpiSCs with TAT-CRE, to excise *LoxP*-flanked *Nanog* exons and bring *GFP* under control of the *Nanog* promoter. This resulted in *Nanog-/-* EpiSCs with *Nanog*-promoter-driven *βgeo* and *GFP* expression. Key: floxed-*Nanog* (F); wild-type *Nanog* (+); deleted *Nanog* (-); endogenous *Nanog* exons (1,2,3,4); splice-acceptor (*SA*); internal ribosome entry site (*IRES*); geneticin resistance (*βgeo*); stop codon (*S*); *LoxP* site (*L*); puromycin resistance (*Pac*); *GFP* coding sequence without a promoter (*GFP*).

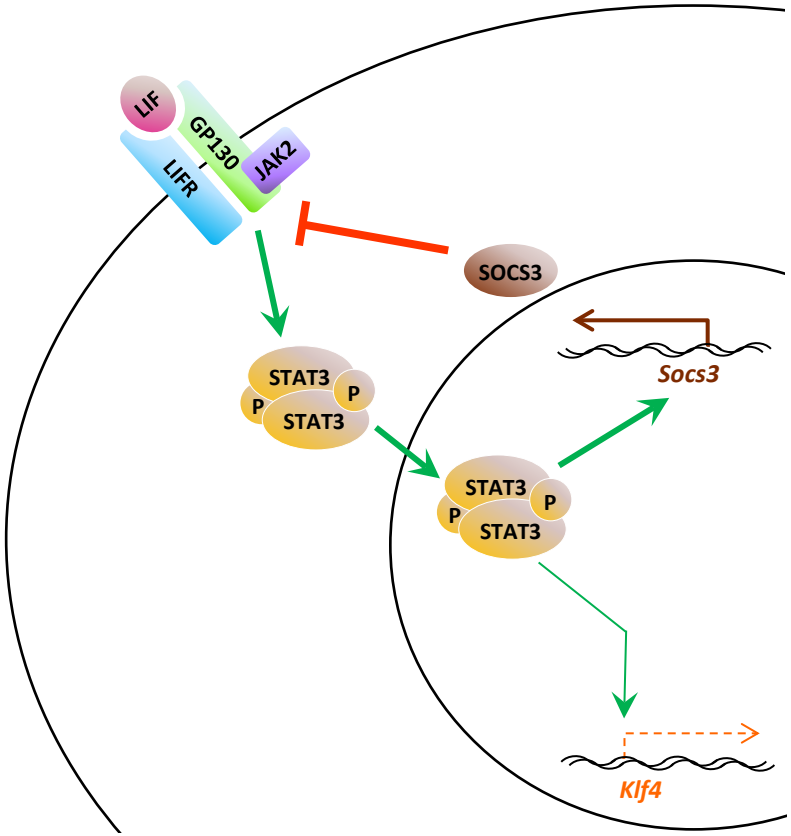
(B) Genotyping PCR was conducted with 3 primers, 1 of which is common to *F*, + and *βgeo* alleles. 1 primer recognizes both *F* and *+* alleles, to give an 800 bp product. 1 primer anneals to the *βgeo* allele, yielding a 600 bp product. L denotes a molecular weight ladder.
Left: Genotyping of embryo-derived EpiSCs revealed that #2 was *Nanog-/F*. Right: *Nanog-/-* clonal EpiSC lines (2cre02, 11, and 15) were obtained after treating #2 Nanog*-/F* EpiSCs with TAT-CRE. *Nanog-/-* ESC samples containing different knock-out alleles were included as negative controls (–), while #2 *Nanog-/F* and #4 *Nanog+/F* EpiSCs provided positive controls. *Nanog-/-* 2cre15 EpiSCs were used in all subsequent experiments.

(C) RT-qPCR analysis of *Nanog* and *Oct4* expression in *Nanog-/-* iNANOG+GY118F EpiSCs following induction with dox (D) and/or GCSF (G) in FGF2+ActivinA. Wild-type ESCs in 2i+LIF (2i+L), and *Nanog-/-* and *Nanog+/F* EpiSCs in FGF2+ActivinA conditions were included as controls. Gene expression was measured relative to *Gapdh* and normalized to the highest value. Data shown are the mean of 3 technical replicates and are from 1 of 2 representative experiments. Error bars indicate ± s.d.

Figure S4:


A
S+L
2i+L
Nanog-/- pre-iPSCs
0
2
7
11d
NANOG↑
Control
CAG-GFP
iPSCs✓
30 ± 6 %
X
C
S+L
Ch+L or KSR+L
2i+L
Nanog-/- pre-iPSCs
0 1
8
20
26d
Ch+L → 2i+L
KSR+L→ 2i+L
CAG-GFP
iPSCs✓
0.005 ± 0.007 %
iPSCs✓
0.14 ± 0.03 %
E
Nanog-/- iPSC mid gestation chimerae
CAG-GFP
B
Relative expression
1
0
WT ESCs 2i+L
Control
NANOG↑
Nanog-/- pre-iPSCs
NANOG↑ Nanog-/- iPSCs 2i+L
Nanog
Endo. Oct4
Rex1
D
Relative expression
2
1
0
WT ESCs 2i+L
Nanog-/- pre-iPSCs
Nanog-/- iPSCs KSR+L
Nanog-/- iPSCs Ch+L
Retro. Oct4
Retro. c-Myc
Retro. Klf4
F
Color Key and Histogram
Count
0
15
0.88
0.92
0.96
1
Value
Nanog-/- ESCs #2
Nanog-/- ESCs #1
Nanog-/- iPSCs (pre) #1
Nanog-/- iPSCs (pre) #2
WT ESCs #2
WT ESCs #1
Nanog-/- iPSCs (Epi) #2
Nanog-/- iPSCs (Epi) #1
Nanog-/- EpiSCs #1
Nanog-/- EpiSCs #2
Nanog-/- pre-iPSCs #2
Nanog-/- pre-iPSCs #1
Nanog-/- pre-iPSCs #1
Nanog-/- pre-iPSCs #2
Nanog-/- EpiSCs #2
Nanog-/- EpiSCs #1
Nanog-/- iPSCs (Epi) #1
Nanog-/- iPSCs (Epi) #2
WT ESCs #1
WT ESCs #2
Nanog-/- iPSCs (pre) #2
Nanog-/- iPSCs (pre) #1
Nanog-/- ESCs #1
Nanog-/- ESCs #2

G


LIF
GP130
JAK2
LIFR
SOCS3
STAT3
P
STAT3
P
P
STAT3
P
P
STAT3
Socs3
Klf4

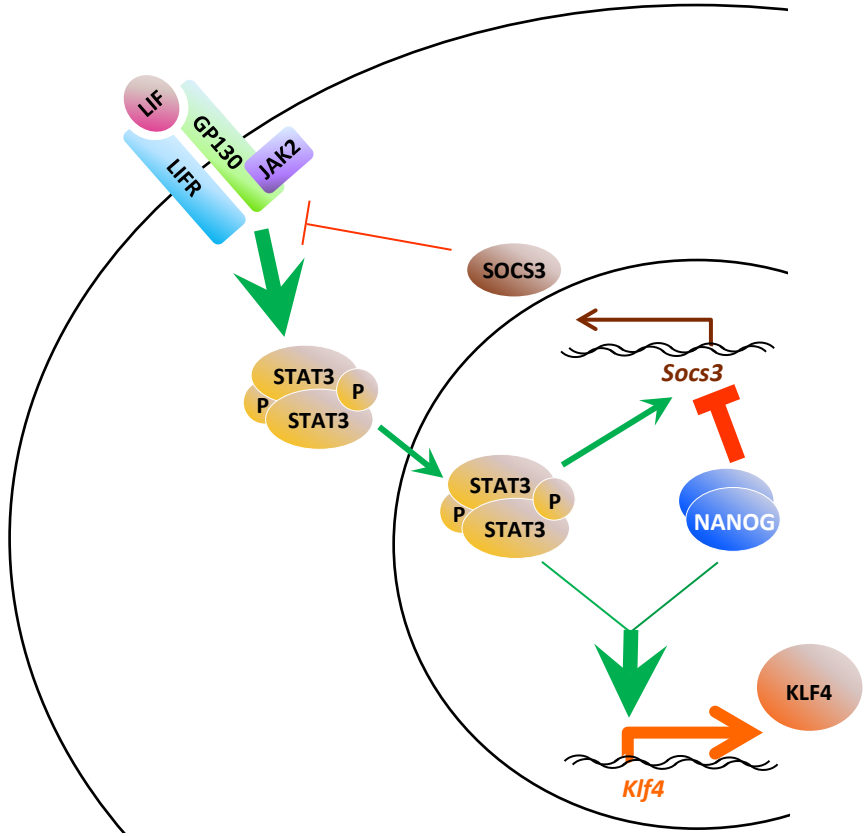

LIF
GP130
JAK2
LIFR
SOCS3
STAT3
P
STAT3
P
P
STAT3
P
P
STAT3
Socs3
NANOG
KLF4
Klf4

Figure S4: Bypass of NANOG in reprogramming (Related to Figure 4)

(A–E) Constitutively GFP+ (*CAG*-GFP) *Nanog-/-* NSCs were derived from chimerae, then converted to pre-iPSCs by retroviral transduction with *Klf4, Oct4* and *c-Myc* in serum+LIF, as previously described [S19]. Pre-iPSCs are non-pluripotent, transgene-dependent, proliferative reprogramming intermediates [S20]. NANOG is dispensable for the initial formation of pre-iPSCs, but is essential for pre-iPSCs to transit to naïve pluripotency in 2i+LIF [S19] (confirmed in A, B). Since *Nanog-/-* EpiSCs were able to reprogram in chiron+LIF but not 2i+LIF (Figure 4), we tested whether *Nanog-/-* pre-iPSCs could also reprogram in alternative conditions.

(A) *Nanog-/-* pre-iPSCs with constitutive NANOG-overexpression or control transgenes were plated in serum+LIF (S+L) and, after 2 days, medium was switched to 2i+LIF (2i+L) to prompt reprogramming. On day 7, G418 was added to select for emergent iPSCs, since *βgeo* is under control of an endogenous *Nanog* promoter. On day 11, representative phase and *CAG*-GFP images were taken and iPSC colonies were counted. The mean number of iPSC colonies is indicated as % of pre-iPSCs initially plated, ± s.d. (n=3 biological replicates).

(B) RT-qPCR analysis of gene expression in 2i+LIF-derived NANOG-rescue iPSCs, compared to parental and control pre-iPSCs in serum+LIF and wild-type ESCs in 2i+LIF. Reprogramming to naïve iPSCs was confirmed by reactivation of endogenous pluripotency genes and retroviral factor silencing. Gene expression was measured relative to *Gapdh* and normalized to the highest value. Data shown are the mean of 3 technical replicates and are from 1 of 2 representative experiments. Error bars indicate ± s.d.

(C) *Nanog-/-* pre-iPSCs were plated in serum+LIF (S+L) and, after 1 day, medium was switched to chiron+LIF (Ch+L) or KSR+LIF (KSR+L) to prompt reprogramming. On day 8, G418 was added to select for emergent iPSCs, since *βgeo* is under control of an endogenous *Nanog* promoter. On day 20, medium was switched to 2i+LIF (2i+L) to further select for naïve iPSCs. On day 26, representative phase and *CAG*-GFP images were taken and iPSC colonies were counted. The mean number of iPSC colonies is indicated as % of pre-iPSCs initially plated, ± s.d. (n=3 biological replicates). We successfully obtained *Nanog-/-* iPSCs in chiron+LIF and KSR+LIF conditions, albeit with low speed and efficiency. Like those generated from EpiSCs, the *Nanog-/-* iPSCs derived from pre-iPSCs could be maintained in 2i+LIF after naïve pluripotency establishment.

(D) RT-qPCR analysis of gene expression in chiron+LIF- and KSR+LIF-derived *Nanog-/-* iPSCs after passaging in 2i+LIF, compared to *Nanog-/-* pre-iPSCs in serum+LIF and wild-type ESCs in 2i+LIF. Reprogramming to naïve iPSCs was confirmed by reactivation of endogenous pluripotency genes and retroviral factor silencing. Gene expression was measured relative to *Gapdh* and normalized to wild-type ESC level. Data shown are the mean of 3 technical replicates and are from 1 of 2 representative experiments. Error bars indicate ± s.d.

(E) Contribution of constitutively GFP+ pre-iPSC-derived *Nanog-/-* iPSCs to mid-gestation chimerae, formally demonstrating their reacquisition of developmental pluripotency.

(F) Correlation heatmap of global gene expression between: wild-type ESCs, *Nanog-/-* ESCs, pre-iPSC-derived *Nanog-/-* iPSCs, pSTAT3+KLF4 EpiSC-derived *Nanog-/-* iPSCs all in 2i+LIF; parental *Nanog-/-* pre-iPSCs in serum+LIF; parental *Nanog-/-* EpiSCs in FGF2+ActivinA. 2 technical replicates were included for each cell line.

(G) Simplified schematic depicting the newfound role of NANOG in the modulation of LIF/STAT3 signal transduction. Left: Existing model of the LIF/STAT3 pathway (see Figure S1A for details). *Klf4* is a known target of active pSTAT3 [S5, S21, S22]. Right: We discovered that NANOG represses *Socs3* and amplifies LIF signal transduction, resulting in higher levels of active pSTAT3. Furthermore, we found that NANOG and pSTAT3 upregulate *Klf4* in a co-dependent and synergistic manner. Functionally, pSTAT3 elevation and KLF4 expression allowed us to bypass the requirement for NANOG in reprogramming.

Supplemental Experimental Procedures

Cell culture

The following established cell lines were used: wild-type E14TG2A ESCs; E14TG2A-derived *Nanog-/-* BT12 and 44cre6 ESCs [S23]; E14TG2A-derived constitutively NANOG-overexpressing RHN ESCs [S24]; *Stat3-/-* and littermate-derived *Stat3+/+* ESCs [S25]; *Nanog-/-* pre-iPSCs [S19]. ESCs and iPSCs were cultured in 2i±LIF or serum+LIF as indicated, pre-iPSCs were cultured in serum+LIF, and EpiSCs were cultured in FGF2+ActivinA. 2i medium [S25] was composed of N2B27, 3 μM CHIR99021 and 1 μM PD0325901 (Stewart lab, Dresden). N2B27 medium comprised 1:1 DMEM/F-12 and Neurobasal (Gibco), 2 mM L-glutamine (Gibco), 1x penicillin-streptomycin (Sigma), 0.1 mM 2-mercaptoethanol (Gibco), 1% B27 (Gibco) and 0.5% N2 (homemade). Serum medium contained GMEM (Sigma), 10% fetal bovine serum (Gibco), 1x non-essential amino acids (Gibco), 1 mM sodium pyruvate (Sigma), 2 mM L-glutamine (Gibco), 1x penicillin-streptomycin (Sigma) and 0.1 mM 2-mercaptoethanol (Gibco). FGF2+ActivinA medium was composed of N2B27, 12.5 ng/ml FGF2 and 20 ng/ml ActivinA (Hyvonen lab, Cambridge). During reprogramming, chiron+LIF and KSR+LIF were used when indicated. Chiron+LIF was the same as N2B27-based 2i+LIF, but without PD0325901. KSR medium contained GMEM (Sigma), 10% KnockOut Serum Replacement (Invitrogen), 1% fetal bovine serum (Gibco), 1x non-essential amino acids (Gibco), 1 mM sodium pyruvate (Sigma), 2 mM L-glutamine (Gibco), 1x penicillin-streptomycin (Sigma) and 0.1 mM 2-mercaptoethanol (Gibco). As required, media were supplemented with 20 ng/ml murine LIF (Hyvonen lab, Cambridge), 30 ng/ml human GCSF (Peprotech), or 1 μg/ml doxycycline (MP Biomedicals). For ESCs, iPSCs and pre-iPSCs, tissue-culture flasks were coated with 0.15% gelatin (Sigma) in PBS (Sigma). For EpiSCs, tissue-culture flasks were coated with 10 μg/ml fibronectin (Millipore) in PBS (Sigma). During expansion of cell lines, selection with 150 μg/ml hygromycin-B (Life Technologies), 20 μg/ml blasticidin (Gibco), 200 μg/ml G418 (Invitrogen) or 100 μg/ml zeocin (Invitrogen) was applied as appropriate. Note that G418 selects for endogenous *Nanog* promoter activity in *Nanog-/-* BT12 and 44cre6 ESCs [S23]. Colonies were dissociated with accutase (Millipore) during passaging.

Cell transfection

For transfections, 1 μg PiggyBac (PB) vector, 2 μg PBase expression vector (*pCAGPBase*) and 10 μl Lipofectamine-2000 (Invitrogen) were incubated for 20 min in 500 ml DMEM (Gibco), then applied to 500,000 cells/10 cm^2 well in 3 ml medium for 24 hours. Selection was applied to transfectants for at least 5 passages prior to use. PB-vectors stably integrate into host genome TTAA-sites in the presence of PBase transposase; *pCAGPBase* does not integrate and is lost with passaging (Cambridge Bioscience). Existing PB vectors were used to drive constitutive expression of: NANOG (*CAG-Nanog-PGK-Hph*); GY118F (*CAG-DsRed-IRES-Hph-CAG-Gy118f*, coding sequence as described in Niwa *et al*, 1998); KLF4 (*CAG-Klf4-IRES-Zeo*); ESRRB (*CAG-Esrrb-IRES-Zeo*)*;* Control (*CAG-empty-PGK-Hph*). A selective reporter for naïve pluripotency was provided when necessary by O4GIP PB vector (*Oct4-distal-enhancer-GFP-IRES-puro*)*.*

Derivation of EpiSCs from post-implantation embryos

EpiSCs were derived from E6.5 embryos carrying the modified alleles shown in Figure S3A. Epiblasts were manually dissected from extra-embryonic tissues and plated on fibronectin-coated plates in FGF2+ActivinA medium. After 5–7 days of culture, regions of the explant exhibiting EpiSC morphology were manually passaged to a fresh plate. Subsequent passages were performed using accutase.

Genotyping

Genomic DNA was extracted by incubation at 95°C for 20 min in 25 mM NaOH + 0.2 mM EDTA, followed by addition of an equal volume of 40 mM Tris HCl and vortexing. Genotyping PCR was conducted on the genomic DNA using Taq DNA polymerase (Qiagen) and the following thermal

cycler settings: 94°C for 5 min, 35 cycles of (94°C for 10 s, 60°C for 20 s, 72°C for 60 s), and 72°C for 3 min. An equal mix of 3 primers was used: βgeo AATGGGCTGACCGCTTCCTC; S5 ACCTCAGCCTCCAGCAGATG; A53 CAGAATGCAGACAGGTCTACAGCCCG. An 800 bp product is amplified from *Flox* and wild-type *Nanog* alleles, whereas the *βgeo* allele yields a 600 bp product.

Blastocyst injection

Chimerae were generated by standard microinjection methodology using host blastocysts of strain C57BL/6. All animal work was performed in accordance with Home Office guidelines and regulations at the University of Cambridge, UK.

Generation and characterization of doxycycline-inducible *Nanog* transgene

TetO-FLBioNanog-UbiP-rtTA3 was amplified from a previously described dox-inducible *Nanog* construct [S26]. *TetO* denotes the operator/promoter and *rtTA3* encodes the dox-controlled transactivator of the 'Tet-On' system [S27], *FLBioNanog* codes for FLAG- and BIOTIN-tagged murine NANOG, and the ubiquitous promoter *UbiP* drives constitutive expression of *rtTA3*. Amplification primers were designed to contain Gateway-compatible *attB* recombination sites: FW GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGAGGTTCTAGACGAGTTTACT; RV GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACCCGGGGAGCATGT. PCR was performed using Phusion Taq polymerase (New England Biolabs, recommended protocol). Successful amplification was verified on a 1% agarose gel, from which *attB1-TetO-FLBioNanog-UbiP-rtTA3-attB2* was purified using a QIAquick gel extraction kit (Qiagen). *TetO-FLBioNanog-UbiP-rtTA3* was cloned into a PiggyBac (PB) vector, to enable efficient and stable integration into the host genome. A Gateway destination PB vector containing *CAG-attR1-ccdB-attR2-IRES-Bsdr* was kindly provided by Yael Costa. To ensure that *FLBioNanog* is expressed only in the presence of dox, the constitutive *CAG* promoter was removed from the destination vector by restriction digestion and re-ligation. 150 ng of the *attB1-TetO-FLBioNanog-UbiP-rtTA3-attB2* insert was used for Gateway cloning, first into a pDonor vector using BP Clonase, then into the PB destination vector by LR Clonase (Invitrogen). Restriction digestion and DNA sequencing confirmed the production of *PB-TetO-FLBioNanog-UbiP-rtTA3-IRES-Bsdr-PB.* This resulting iNANOG construct confers constitutive blasticidin resistance (*Bsdr*) driven by *UbiP*, allowing selection of successfully transfected cells irrespective of dox presence. 1 μg/ml dox was found to be the optimal concentration for NANOG induction (data not shown). In timed induction experiments, dox was added 2 hours before GCSF, to account for the lag between transcription and translation of iNANOG (Figure S1D).

Western blotting

Dissociated cells were lysed in RIPA buffer (as described by Sigma) containing Complete-ULTRA protease-inhibitor and PhoStop phosphatase-inhibitor cocktails (Roche), and sonicated with Bioruptor200 (Diagenode) at high frequency, alternating 30 s on/off for 3 min. If necessary, lysate protein concentration was measured by BCA microplate procedure (Pierce) and adjusted prior to loading. SDS-PAGE electrophoresis was performed using NuPAGE 10% Bis-Tris gels (Invitrogen), then proteins were transferred onto a nitrocellulose membrane (Amersham) at 395 mA for 70 minutes. Membranes were stripped between pSTAT3 and total STAT3 blots (stripping buffer and protocol as described by Abcam). The following primary antibodies and dilutions were used: rabbit monoclonal against p-Y705-STAT3 (9145 1:1000) and rabbit polyclonal against total STAT3 (9132 1:1000) from Cell Signalling Technology; mouse monoclonal against OCT4 (sc-5279 1:500) from Santa Cruz Biotechnology; mouse monoclonal against α-TUBULIN (AB7291 1:5000) from Abcam; rabbit polyclonal against NANOG (A300-397A 1:5000) from Bethyl Laboratories; goat polyclonal against KLF4 (AF3158 1:1000) from R&D Systems. Detection was achieved using HRP-linked secondary antibodies against the appropriate species from GE Healthcare (anti-rabbit, anti-rat, anti-mouse) and Santa Cruz Biotechnology (anti-goat), and ECL Plus Western Blotting Detection System (GE Healthcare).

RT-qPCR

Total RNA was extracted using RNeasy kits, according to manufacturer's spin protocol, including on-column DNaseI digest (Qiagen). cDNA was produced from 1 µg RNA using SuperscriptIII VILO cDNA synthesis kit, following the recommended protocol including RNaseH treatment (Invitrogen). RT-qPCR reactions were performed in triplicate, using StepOnePlus Real Time PCR System with recommended thermocycler settings (Applied Biosystems) and TaqMan Fast Universal PCR Master Mix (Applied Biosystems). Average gene expression relative to *Gapdh* was determined using FAM-labelled TaqMan assay probe together with VIC-labelled *Gapdh* probe (Applied Biosystems).

Applied Biosystems TaqMan RT-qPCR custom assays:

Gene	**Forward primer (5'–3')**	**Reverse primer (5'–3')**	**Probe (FAM 5'–3' MGB)**
K4eRNA	AGGCTTTGGCTGGCTGATAA	CTGTCTCCATAGGTACTGACTTCCT	CCCAGCTCAGTAATTG
Endogenous *Oct4*	TTCCACCAGGCCCCC	GGTGAGAAGGCGAAGTCTGAAG	CCCACCTTCCCCATGGCT
Retroviral *Oct4*	TGGTACGGGAAATCACAAGTTTGTA	GGTGAGAAGGCGAAGTCTGAAG	CACCTTCCCCATGGCTG
Retroviral *Klf4*	TGGTACGGGAAATCACAAGTTTGTA	GAGCAGAGCGTCGCTGA	CCCCTTCACCATGGCTG
Retroviral *c-Myc*	TGGTACGGGAAATCACAAGTTTGTA	GGTCATAGTTCCTGTTGGTGAAGTT	CCCTTCACCATGCCCC

Applied Biosystems TaqMan RT-qPCR standard assays:

Gene	**Probe ID**
Gapdh	4352339E
Esrrb	Mm00442411_m1
Fgf5	Mm00438919_m1
Klf2	Mm01244979_g1
Klf4	Mm00516104_m1
Klf5	Mm00456521_m1
Nanog	Mm02384862_g1
Nr0b1	Mm00431729_m1
Nr5a2	Mm00446088_m1
Rex1	Mm03053975_g1
Socs3	Mm01249143_g1
Tfcp2l1	Mm00470119_m1

ChIP

NANOG ChIP was performed as follows: cells (10×10^6 for each sample) were fixed for 10 min in 1% formaldehyde, washed with ice-cold PBS and incubated for 10 min in lysis buffer 1 (50 mM HEPES at pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP40 and 0.25% Triton X-100) and then for 10 min in lysis buffer 2 (10 mM Tris at pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA). Nuclei were pelleted, resuspended in shearing buffer (1% SDS, 10 mM EDTA and 50 mM Tris at pH 8.0) and sonicated to obtain an average DNA fragment size of 500 base pairs. Lysates were diluted 1:10 in dilution buffer (50 mM Tris-HCl at pH 8.0, 167 mM NaCl, 1.1% Triton X-100 and 0.11% Na deoxycholate) and pre-cleared for 2 h at 4°C with Dynabeads Protein G magnetic beads (Life Technologies) that were pre-incubated with isotype IgG antibody. The chromatin was then incubated overnight at 4°C with 2 µg of rabbit polyclonal antibody against NANOG (Bethyl Laboratories, A300-397A) or an isotype IgG control (Santa Cruz Biotechnology, sc-2027). Lysates were then incubated for 1 h at 4°C with blocked Dynabeads magnetic beads, and the beads were washed twice in wash buffer 1 (50 mM Tris-HCl at pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na deoxycholate and 0.5 mM EGTA), once in wash buffer 2 (50 mM Tris-HCl at pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na deoxycholate and 0.5 mM EGTA), once in wash buffer 3 (50 mM Tris at pH 8.0, 250 mM LiCl, 0.5% Na deoxycholate, 0.5% NP40, 1 mM EDTA and 0.5 mM EGTA) and twice in wash buffer 4 (50 mM Tris at pH 8.0, 10 mM EDTA and 5 mM EGTA). Chromatin was eluted for 30 min at 65°C in elution buffer (1% SDS and 0.1 M $NaHCO_3$). Samples were incubated overnight at 65°C to reverse the crosslinking and purified using the QIAquick PCR Purification kit (Qiagen). Chromatin was analyzed by Fast SYBR Green RT-qPCR (Applied Biosystems). Enrichment was calculated relative to the Input sample.

ChIP primers:

Region	**Forward primer (5'–3')**	**Reverse primer (5'–3')**
Socs3 Proximal	GAAAAGGCTTGAGGGTCGGA	CGGGCCTGGAATGTCAAACT
Socs3 Distal	TCAGGAGTCCCTGTGCTCTAA	GGCAGACGGGTCTACTTTGAA
Negative control region	CTGGGCTTGCAGCTTAGG	AGAGACCTGGCTGAGGATGAC

Analysis of ChIP-seq datasets

The following published ChIP-seq datasets were used to examine transcription factor binding and histone modifications at genomic regions of interest in ESCs: NANOG (accession number GSE11724); pSTAT3 (GSE11431); MED12 (GSE22562); RNA PolII (GSE23943); H3K27ac (GSE24164); H3K36me3 (GSE23943); H3K4me3 (GSE23943) and H3K27me3 (GSE23943). Raw ChIP-seq datasets were downloaded from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/), realigned to the mouse mm9 genome with Bowtie software (http://bowtie-bio.sourceforge.net), and peaks were called with MACS software (http://liulab.dfci.harvard.edu/MACS) using default parameters. 'K4-K36' domains were predicted from H3K4me3 and H3K36me3 patterns by Guttman *et al,* 2009.

Microarray

Amplification and labelling of RNA were performed using the TotalPrep-96 RNA Amplification Kit for the Illumina platform (Ambion). Subsequent hybridization, staining and scanning were performed according to the Whole Genome Gene Expression Direct Hybridization Guide on the MouseWG-6 v2.0 Expression BeadChip (Illumina). Data were loaded into the R package lumi [S28] and then divided into subsets to be analyzed. The data were transformed using Variance Stabilization [S29] and normalized using quantile normalization. Comparisons were performed in the R package limma [S30] and the results were corrected using the False Discovery Rate. Our analysis employed a 5% confidence interval.

Supplemental References

S1. Yoshida, K., Chambers, I., Nichols, J., Smith, A., Saito, M., Yasukawa, K., Shoyab, M., Taga, T., and Kishimoto, T. (1994). Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. Mechanisms of development *45*, 163-171.

S2. Boeuf, H., Hauss, C., Graeve, F.D., Baran, N., and Kedinger, C. (1997). Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. J Cell Biol *138*, 1207-1217.

S3. Zhong, Z., Wen, Z., and Darnell, J.E., Jr. (1994). Stat3: a STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and interleukin-6. Science *264*, 95-98.

S4. Burdon, T., Smith, A., and Savatier, P. (2002). Signalling, cell cycle and pluripotency in embryonic stem cells. Trends Cell Biol *12*, 432-438.

S5. Bourillot, P.Y., Aksoy, I., Schreiber, V., Wianny, F., Schulz, H., Hummel, O., Hubner, N., and Savatier, P. (2009). Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. Stem cells *27*, 1760-1771.

S6. Kershaw, N.J., Murphy, J.M., Liau, N.P., Varghese, L.N., Laktyushin, A., Whitlock, E.L., Lucet, I.S., Nicola, N.A., and Babon, J.J. (2013). SOCS3 binds specific receptor-JAK complexes to control cytokine signaling by direct kinase inhibition. Nat Struct Mol Biol *20*, 469-476.

S7. Schmitz, J., Weissenbach, M., Haan, S., Heinrich, P.C., and Schaper, F. (2000). SOCS3 exerts its inhibitory function on interleukin-6 signal transduction through the SHP2 recruitment site of gp130. The Journal of biological chemistry *275*, 12848-12856.

S8. Yoshiura, S., Ohtsuka, T., Takenaka, Y., Nagahara, H., Yoshikawa, K., and Kageyama, R. (2007). Ultradian oscillations of Stat, Smad, and Hes1 expression in response to serum. Proceedings of the National Academy of Sciences of the United States of America *104*, 11292-11297.

S9. Fukada, T., Hibi, M., Yamanaka, Y., Takahashi-Tezuka, M., Fujitani, Y., Yamaguchi, T., Nakajima, K., and Hirano, T. (1996). Two signals are necessary for cell proliferation induced by a cytokine receptor gp130: involvement of STAT3 in anti-apoptosis. Immunity *5*, 449-460.

S10. Takahashi-Tezuka, M., Yoshida, Y., Fukada, T., Ohtani, T., Yamanaka, Y., Nishida, K., Nakajima, K., Hibi, M., and Hirano, T. (1998). Gab1 acts as an adapter molecule linking the cytokine receptor gp130 to ERK mitogen-activated protein kinase. Molecular and cellular biology *18*, 4109-4117.

S11. Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes & development *12*, 2048-2060.

S12. Burdon, T., Chambers, I., Stracey, C., Niwa, H., and Smith, A. (1999). Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. Cells, tissues, organs *165*, 131-143.

S13. Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Developmental biology *210*, 30-43.

S14. Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.

S15. Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227.

S16. Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. The EMBO journal *32*, 2561-2574.

S17. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

S18. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

S19. Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722-737.

S20. Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS biology *6*, e253.

S21. Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118-122.

S22. van Oosten, A.L., Costa, Y., Smith, A., and Silva, J.C. (2012). JAK/STAT3 signalling is sufficient and dominant over antagonistic cues for the establishment of naive pluripotency. Nat Commun *3*, 817.
S23. Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230-1234.
S24. Silva, J., Chambers, I., Pollard, S., and Smith, A. (2006). Nanog promotes transfer of pluripotency after cell fusion. Nature *441*, 997-1001.
S25. Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.
S26. Fidalgo, M., Faiola, F., Pereira, C.F., Ding, J., Saunders, A., Gingold, J., Schaniel, C., Lemischka, I.R., Silva, J.C., and Wang, J. (2012). Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. Proceedings of the National Academy of Sciences of the United States of America *109*, 16202-16207.
S27. Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science *268*, 1766-1769.
S28. Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics *24*, 1547-1548.
S29. Lin, S.M., Du, P., Huber, W., and Kibbe, W.A. (2008). Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic acids research *36*, e11.
S30. Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.